Study of the performance of the DAMPE BGO calorimeter

Relatori
Dott. Fabio Gargano
Dott. Francesco Loparco

Laureando
Simone Garrappa

Anno Accademico 2015–2016
## Contents

1. Dark Matter .......................... 6
   1.1 Historical evidences for Dark Matter ......................... 6
   1.2 Candidates for Dark Matter ............................... 9
      1.2.1 Axions ........................................ 9
      1.2.2 Sterile neutrinos ............................... 10
      1.2.3 WIMPs ........................................ 11
   1.3 Indirect Dark Matter detection .............................. 12
      1.3.1 Dark Matter annihilation and decay signals ............. 13
      1.3.2 Search for gamma-ray lines ........................ 16
      1.3.3 Dark Matter annihilation and decay in cosmic ray fluxes 17

2. The DArk Matter Particle Explorer .................. 22
   2.1 The DAMPE mission .................................... 22
   2.2 The DAMPE Detector ................................... 23
      2.2.1 The Plastic Scintillator strip Detector (PSD) ........ 24
      2.2.2 The Silicon-Tungsten Tracker (STK) ................... 27
      2.2.3 The BGO Calorimeter .............................. 30
      2.2.4 The NeUtron Detector (NUD) ......................... 32

3. Beam Test analysis on BGO calorimeter .......... 35
   3.1 Profiles of electromagnetic cascades ........................ 37
3.2 Energy correction .......................................... 40

4. The $e^-/p$ separation multivariate analysis on DAMPE BGO calorimeter .................................. 48
   4.1 The imaging calorimeter approach with Beam Test data ........................................... 50
      4.1.1 Data selection from beam test campaign ......................................................... 51
      4.1.2 Supervised dimensionality reduction via Linear Discriminant Analysis .................. 53
      4.1.3 Support Vector Machines (SVMs) ................................................................. 56
   4.2 Deep Learning classification approach with on-orbit data ...................................... 61
      4.2.1 Monte Carlo events generation and pre-selection ............................................. 62
      4.2.2 The Random Forest Classifier ........................................................................ 65
      4.2.3 The Convolutional Neural Network (ConvNET) .............................................. 72
      4.2.4 A meta-classifier with Random Forest and ConvNET ............................. 80
      4.2.5 First application on real observed data ......................................................... 82

References .................................................................................................................. 87

Conclusions ............................................................................................................... 87

Appendix A. Electromagnetic cascades parameterization ........................................... 89

Bibliography ............................................................................................................. 91
Abstract

The DArk Matter Particle Explorer (DAMPE) is an orbital experiment aiming to search for dark matter by indirect detection, measuring the spectra of gamma-rays, electrons and positrons originating from deep space. The BGO electromagnetic calorimeter is one of the key sub-detectors of DAMPE, which is designed to perform measurements in a wide energy range from 5 GeV up to 10 TeV.

The purpose of this work is to characterize the BGO calorimeter, studying its performance during the beam test campaign, and to develop a machine learning algorithm for particle identification with the calorimeter, in order to estimate its discrimination power is separating the electromagnetic component of the cosmic rays from the hadronic, with a first application on real observation made by DAMPE in its first months on-orbit.

This thesis work consists of four chapters:

• In the first chapter, the most recent models and techniques of indirect dark matter search are described, along with results from other experiments in this field.

• In the second chapter the DAMPE detector is described, focusing on structure and proprieties of each sub-detector.

• In the third chapter a characterization analysis on beam test data with electrons is performed, with the application of correction methods for
the reconstructed energy inside the calorimeter and the characterization of its energy resolution and linearity.

• In last chapter, the machine learning algorithm is developed using the Monte Carlo data productions to train the classifier that is then used to estimate the amount of different cosmic rays components on observed data.
Chapter 1

Dark Matter

1.1 Historical evidences for Dark Matter

The existence of Dark Matter (DM) is nowadays well established, thanks to several convincing evidences given by cosmology and astrophysics. The first evidence for DM came from observation that various celestial objects (gas clouds, stars, globular clusters or galaxies) move faster than expected when considering only the effects of the gravitational attraction of other visible objects. The most important historical example comes from the measurement of galactic rotation curves, started with the contributions of Jan Hendrik Oort (1932) and Horace Babcock (1939), and culminated with the discovery by Vera Rubin and Kent Ford that most stars in spiral galaxies orbit at roughly the same speed. The rotational velocity $v$ of an object on a stable Keplerian orbit with radius $r$ around a galaxy scales as $v(r) \propto \sqrt{M(r)/r}$ where $M(r)$ is the mass inside the orbit. If $r$ lies outside the visible part of the galaxy and is tracked by light, one would expect $v(r) \propto 1/\sqrt{r}$. Instead, in most of galaxies one finds that $v$ is approximately constant up to the largest value of $r$ where the rotation curve can be measured. In the neighborhood of our Solar System, $v \sim 240$ km/s with small variations up to the largest observable radius in our galaxy. This implies the existence of a dark halo,
with mass density $\rho(r) \propto 1/r^2$ and, in order to keep the mass of the galaxy finite, at some point $\rho$ must drop faster [2]. A good approximation in the central regions of halos is to model them as spherically-symmetric distributions. In the late 1990s simulations showed that the Navarro-Frenk-White (NFW) 2-parameter model described with high accuracy the density profile of dark matter halos over a wide range of halo-masses:

$$\rho_{\text{NFW}}(r) = \frac{\rho_0}{\left(\frac{r}{r_s}\right)^2 \left[1 + \left(\frac{r}{r_s}\right)^2\right]^\frac{1}{2}} \tag{1.1}$$

where $r$ is the distance from the center of the halo and $r_s$ is a scale radius that for the Milky Way is $\sim 20\text{kpc}$. Recent simulations show that the dark matter density profile in the innermost regions of the halo exhibits deviations from a simple power law, and can be better fitted with the profile proposed by Einasto:

$$\rho_{\text{Ein}}(r) = \rho_0 \exp\left\{-\frac{2}{a} \left[\left(\frac{r}{r_s}\right)^a - 1\right]\right\}. \tag{1.2}$$

This profile introduces an extra shape parameter $\alpha(\sim 2$ for the Milky Way) with respect to the standard NFW profile.

Observations of dwarf spheroidal and low-surface-brightness galaxies have shown that some objects are better described by the Burkert profile:

$$\rho_{\text{Burk}}(r) = \frac{\rho_0}{\left(1 + \frac{r}{r_s}\right)^2 \left(1 + \frac{r^2}{r_s^2}\right)} \tag{1.3}$$

This profile better describes objects with a central core and exhibits constant density for radii much smaller than the scale radius ($r_s \sim 6\text{kpc}$ for the Milky Way).

The three density profiles are illustrated in Fig 1.1 using the value of the parameters for the Milky Way. The distributions are very similar outside
the solar circle ($r \sim 8$ kpc), but can have substantially different behaviors in the inner regions of the halo, that result in large variations in indirect dark matter signals. N-body simulations in $\Lambda$CDM (Lambda Cold Dark Matter) cosmologies show that dark matter halos are populated with smaller, denser halos, called subhalos of mass $\sim 10^{-6} M_\odot$ (where $M_\odot$ is the solar mass) or smaller.

The most recent estimates [4] find a dark matter density in the neighborhood of our solar system of:

$$\rho_{DM}^{local} = (0.39 \pm 0.03) \frac{GeV}{cm^3}$$  \hspace{1cm} (1.4)

This value could be increased by a factor $(1.2 \pm 0.2)$ since the baryons in the galactic disk also increase the local DM density. The small subhalo structures are not likely to change the local DM density significantly.
1.2 Candidates for Dark Matter

Studies on the structure formation in the Universe indicate that most DM should have been non-relativistic at the beginning of galaxy formation. For this reason it is often said that DM should be "cold". Currently the most accurate determination of the dark matter density $\Omega_{DM}$ (where $\Omega_X \equiv \rho_X/\rho_{crit}$, $\rho_{crit}$ being the critical mass density) comes from global fits of cosmological parameters to a variety of observations. For example, the density of cold, non-baryonic matter found using measurements of the anisotropy of the cosmic microwave background (CMB) and of spatial distribution of galaxies is

$$\Omega_{nbdm}h^2 = 0.1198 \pm 0.0026,$$

(1.5)

where $h$ is the Hubble constant in units of 100 km/(s·Mpc). Candidates for this non-baryonic DM must satisfy some peculiar conditions:

- they must be stable on cosmological time scales (otherwise they would have been decayed by now);
- they must interact very weakly with electromagnetic radiation ;
- they must have the correct relic density;

Candidates satisfying these conditions include axions, sterile neutrinos and weakly interactive massive particles (WIMPs).

1.2.1 Axions

The axions are pseudo Nambu-Goldstone bosons associated with the spontaneous breaking of a new global "Peccei-Quinn" (PQ) U(1) symmetry at scale $f_a$. Their existence was first postulated to solve the strong CP
problem of QCD, and they also occur naturally in superstring theories. Although very light, axions are valuable candidates for cold DM, since they were produced non-thermally. At temperatures well above the QCD phase transition, the axion is massless, and the axion field can take any value, parameterized by the "misalignment angle" $\Theta_i$. At $T \lesssim 1$ GeV, the axion develops a mass $m_a \sim f_\pi m_\pi / f_a$ due to instanton effects. Unless the axion field happens to find itself at the minimum of its potential ($\Theta_i = 0$), it will begin to oscillate once $m_a$ becomes comparable to the Hubble parameter $H$. These coherent oscillations transform the energy originally stored in the axion field into physical axion quanta. The contribution of this mechanism to the present axion relic density is

$$\Omega_a h^2 = \kappa_a \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{1.175} \Theta_i^2$$ \hspace{1cm} (1.6)

where the numerical factor $\kappa_a$ lies roughly between 0.5 and a few units. If $\Theta_i \sim O(1)$, Eq. 1.6 will saturate Eq. 1.5 for $f_a \sim 10^{11}$ GeV. This would correspond to an axion mass of $\sim 0.1$ meV [2].

### 1.2.2 Sterile neutrinos

The neutrino candidate for dark matter is right-handed and is often referred as "sterile". In general, the neutrino flavor eigenstates ($\nu_\alpha$, with $\alpha = e, \mu, \tau$) are a linear combination of mass eigenstates ($\nu_a$, with $a = 1, 2, \ldots$) and the sterile neutrino has a very small mixing with active neutrinos. A heavier mass state can radiatively decay to lighter mass state ($\nu_2 \rightarrow \nu_1 + \gamma$, with $m_2 > m_1$), and since the sterile neutrino is predominantly composed of $\nu_2$, this process is often described as the sterile neutrino decaying to an active neutrino. This produces a photon line at half of the sterile neutrino mass, which for most viable dark matter candidates is in the keV to MeV
1.2 Candidates for Dark Matter

energy range. Line emission provides therefore a means to indirectly detect sterile neutrino dark matter[3].

1.2.3 WIMPs

Weakly interactive massive particles (WIMPs) $\chi$ are particles with mass roughly between 10 GeV and a few TeV, having decay and annihilation cross-sections with values similar to the typical electroweak cross-sections. Their present relic density can be calculated assuming that they were in thermal and chemical equilibrium with the hot "soup" of Standard Model (SM) particles after cosmic inflation. Following the Boltzmann distribution, their density would become exponentially suppressed at $T < m_\chi$. The WIMPs therefore drop out of thermal equilibrium ("freeze out") once the rate of reactions that change SM particles into WIMPs or vice versa (which is proportional to the product of the WIMP number density and the WIMP pair annihilation cross section into SM particles $\sigma_A$ times the WIMP velocity) becomes smaller than the Hubble expansion rate of the Universe. After freeze out, the co-moving WIMP density remains essentially constant; if the Universe evolved adiabatically after WIMP decoupling, this implies a constant WIMP number to entropy density ratio. Their present relic density is then approximately given by [6]:

$$\Omega_\chi h^2 \simeq \text{const.} \cdot \frac{T_0^3}{M_{Pl}^2 (\sigma_A v)} \simeq \frac{0.1pb \cdot c}{(\sigma_A v)}. \quad (1.7)$$

Here $T_0$ is the current CMB temperature, $M_{Pl}$ is the Planck mass, $c$ is the speed of light, $\sigma_A$ is the total annihilation cross section of a pair of WIMPs into SM particles, $v$ is the relative velocity between two WIMPs in their center-of-mass system, and $\langle ... \rangle$ denotes thermal averaging. Freeze out happens at temperature $T_F \simeq m_\chi/20$ almost independently of the properties of the WIMP. This means that WIMPs are already non-relativistic when they
1.3 Indirect Dark Matter detection

decouple from the thermal plasma. It is interesting to see that the term 0.1 pb in the r.h.s. of Eq. \[ \text{contains the factors } T_0 \text{ and } M_{Pl}, \] and results to be near the typical size of weak interaction cross section.

After all these considerations, the most likely WIMP candidate is a heavy neutrino. However, a SU(2) doublet neutrino will have a too small relic density if its mass exceeds \( M_Z/2 \), as required by the LEP data. One can suppress the annihilation cross section, and hence increase the relic density, by postulating a mixing between a heavy SU(2) doublet and some sterile neutrino that must also be stable.

The currently best motivated WIMP candidate is the lightest superparticle (LSP) in supersymmetric models with exact R-parity, which guarantees its stability. Searches for exotic isotopes imply that a stable LSP has to be neutral. This leaves basically two candidates among the superpartners of ordinary particles: a sneutrino and a neutralino. The negative outcome of various WIMP searches rules out "ordinary" sneutrinos as primary component of the DM halo of our galaxy. The most widely studied WIMP is therefore the lightest neutralino. Detailed calculations show that the lightest neutralino will have the desired thermal relic density of Eq. \[ \text{in at least four distinct regions of the parameter space. The particle } \chi \text{ could be a bino or photino (the superpartners of the } U(1)_Y \text{ gauge boson and photon, respectively), if both } \chi \text{ and some sleptons have mass below } \sim 150 \text{ GeV, or if } m_\chi \text{ is close to the mass of some sfermion, or if } 2m_\chi \text{ is close to the mass of the } CP\text{-odd Higgs boson present in supersymmetric models} \] \[.\]

1.3 Indirect Dark Matter detection

Indirect dark matter detection is a technique that uses astronomical observations of Standard Model (SM) particles to detect the products of the
annihilation or decay of dark matter. It is distinguished from direct detection, which is based on the detection of scattering dark matter particles with nuclei in laboratory.

Indirect dark matter detection approaches offer the advantage of being able to identify dark matter particles in an astrophysical environment, providing an independent tool to map the dark matter distribution and also reveal and understand the detailed structure of dark matter halos.

The potential of indirect detection can be further enhanced by leveraging the full complementary of direct and collider approaches. Moreover, to remain strictly in the investigation field of the DAMPE detector, only the main research topics in indirect detection will be now reviewed, and the state-of-the-art research and the results of space experiments will be shown, in particular from the Fermi-LAT and AMS-02 missions.

### 1.3.1 Dark Matter annihilation and decay signals

Dark Matter annihilation is the main interaction to which gamma-ray space-experiments like Fermi and DAMPE are sensitive. Gamma rays can be produced by DM annihilation in two major ways:

- annihilation into other particles, which eventually yield gamma rays either through $\pi^0$ production or final state bremsstrahlung and Inverse Compton radiation from leptonic channels. This will result in a continuous gamma-ray energy spectrum;

- direct annihilation to $\gamma X$, where $X$ is usually another neutral state, typically a photon, a $Z$ or a Higgs boson. This will result into a spectral line.

The expected DM-induced gamma-ray flux from a region covering a solid angle $\Delta\Omega$ and centered on a DM source can be factorized as:
\[ \Phi_\gamma(E, \Delta\Omega) = J(\Delta\Omega) \times \Phi^{PP}(E) \] (1.8)

where \( J(\Delta\Omega) \) is "the astrophysical factor" (or "J-factor") defined as:

\[ J(\Delta\Omega) = \int_{\Delta\Omega} d\Omega \int_{\text{l.o.s.}} dl \rho^2(l(\Omega)) \] (1.9)

i.e. it is the line-of-sight integral of the squared DM density.

The factor \( \Phi^{PP}(E) \) is the "particle-physics factor", that for a given value \( m_\chi \) of the DM mass is given by:

\[ \Phi^{PP}(E) = \frac{1}{4\pi} \frac{\langle \sigma_A v \rangle}{2m_\chi^2} \sum_f N_f(E, m_\chi) B_f. \] (1.10)

In this equation \( \langle \sigma_A v \rangle \) indicates the velocity averaged annihilation cross section, while \( B_f \) and \( N_f(E, m_\chi) \) are respectively the branching ratio and the differential photon spectrum of each pair annihilation final state \( f \).

In the case of DM decay, the production spectrum can be obtained from the previous equations replacing \( \langle \sigma_A v \rangle/2m_\chi^2 \) with \( \Gamma/m_\chi \), where \( \Gamma \) is the DM decay rate, and \( \rho^2 \) with \( \rho \).

The best targets to search for these annihilation signals must be regions with high DM density and low background. Table 1.1 summarizes the J-factors for some selected targets [3]. The Galactic Center has the largest J-factor, but strong astrophysical backgrounds represent a limit for dark matter search. Satellite galaxies tend to provide cleaner targets and a joint analysis of multiple satellite galaxies can help to compensate for the lower J-factors of individual ones. Cluster of galaxies appear to be less promising targets; however the uncertainty on the J-factor due to substructures is quite large, and substructures could significantly enhance the signal at large distances from the cluster center. Milky Way DM subhalos that do not host a luminous compo-
1.3 Indirect Dark Matter detection

...nent could be detected either as individual sources, or their collective signal could contribute to the measured diffuse emission.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|}
\hline
Target & \(\log_{10}(J_{\text{ann}})\) \\
\hline
Galactic Center & 21.5 \\
Dwarf Galaxies (best) & 19 \\
Galaxy Clusters (best) & 18 \\
\hline
\end{tabular}
\caption{Approximate J-factors for selected targets, integrated over a circular region with angular radius of 0.5\(^\circ\), with \(J_{\text{ann}}\) in \(GeV^2 cm^{-5} sr\) from [3]}
\end{table}

Milky Way dwarf spheroidal (dSph) galaxies are probably the most promising target regions for DM searches because their high mass to light ratio is predicted to be from 100 to 1000, implying that they could be highly DM-dominated systems [7] [8]. The high Galactic latitudes where dSph galaxies are located represent the greatest advantage for these object, since they have low astrophysical background at such Galactic latitudes and this makes dSph galaxies an optimal target for DM searches.

The Fermi-LAT did not detect significant emission from any of such objects, but the upper limits on the gamma-ray flux allowed to set very tight constraints on the DM annihilation cross section. The analysis of a set of 15 known dSph galaxies using 6 years of LAT data [9] has allowed to set upper limits on the cross section in several annihilation channels. As an example, Fig. 1.2 shows the 95\% confidence level upper limit for the annihilation cross section in the \(b\bar{b}\) channel as a function of the WIMP mass. The limits are consistent with the expected values in case of signal absence and are well below the thermal relic production cross section for WIMP masses up to 100 GeV. Similar results are found for other annihilation channels.
1.3 Indirect Dark Matter detection

Figure 1.2: Constraints on the DM annihilation cross section at 95% confidence level for the $b\bar{b}$ channel derived from a combined analysis of 15 dSphs. Bands for the expected sensitivity are calculated by repeating the same analysis on 300 randomly selected sets of high-Galactic-latitude blank fields. The dashed line shows the median expected sensitivity while the bands represent the 68% and the 95% quantiles. The solid blue curve shows the limits derived from a previous analysis of four years of LAT data and the same sample of 15 dSph galaxies. The plot is reprinted from ref. [10].

1.3.2 Search for gamma-ray lines

A line at the WIMP mass, due to for instance to the $\gamma\gamma, \gamma Z^0$ or $\gamma H^0$ production channel, could be observed as a feature in the astrophysical source spectrum. Such an observation in the high energy region would be a smoking gun for WIMP DM as it cannot be explained by any standard astrophysical mechanism of $\gamma$ emission.

A gamma-ray line from DM annihilation in the channel $\gamma X$ would be found at the energy:

$$E_\gamma = m_X \left( 1 - \frac{m_X^2}{4m_X^2} \right)$$  \hspace{1cm} (1.11)
where $m_\chi$ is the WIMP mass and $m_X$ is the mass of the particle produced with the photon. For decays this energy value should be changed replacing $m_\chi$ with $m_\chi/2$ and, assuming that the typical velocity of $\chi$ is very small (of the order of $v/c \sim 10^{-3}$) these signals should be approximately monochromatic in the lab frame as well.

The most recent LAT data analysis has been performed using a 5.8 years data sample in the energy range from 200 MeV to 500 GeV [10]. The line search has been performed looking at the data from a set of five regions of interest (ROIs) optimized for sensitivity to WIMP annihilation or decay and for different DM distribution profiles in the Galaxy: NFW, an adiabatically contracted NFW (NFWc), Einasto, and a cored isothermal profile. Fig. 1.3 shows the upper limits at 95% confidence level on the velocity averaged annihilation cross section $\langle \sigma_A v \rangle_{\gamma\gamma}$ for the different DM profiles, each considered in the corresponding optimized ROI. No evidence of gamma-ray lines is found in the data.

1.3.3 Dark Matter annihilation and decay in cosmic ray fluxes

A unique probe of local DM annihilation and decay is represented by the measurements of cosmic-ray fluxes at Earth. Cosmic-ray (CR) observations complete the information given by photon and neutrino searches, and can place strong constraints on branching ratios to specific channels. Several experiments have observed that the positron fraction $(e^+/(e^+ + e^-))$ rises from $\sim 10$ GeV to at least $\sim 100$ GeV. The most important results have been achieved from PAMELA [11], the Fermi LAT [12] and most recently by AMS-02 [13] (Fig. 1.4).

The rise in the positron fraction does not agree with the expectations of
Figure 1.3: Upper limits at 95% confidence level on the velocity averaged annihilation cross section \( \langle \sigma A v \rangle_{\gamma\gamma} \) for four different DM profiles, each considered in the corresponding optimized ROI. Yellow (green) bands show the 68% (95%) expected containments derived from 1000 Monte Carlo simulations without DM. The black dashed lines show the median expected limits from those simulations. Also shown are the limits obtained in previous 3.7-year and 5.2-year line searches, when the assumed DM profiles were the same. The plot is reprinted from ref. [10].

Positrons originated by interactions of Galactic cosmic rays, which should result into a drop in the positron fraction above \( \sim 10 \text{ GeV} \). The observed rise suggests that there must be a source of high energy positrons in the neighborhood. Dark matter annihilation or decay is one possible candidate, although the presence of pulsars have also been proposed. Both candidates are expected to inject \( e^+ - e^- \) pairs, and are sufficiently close to the Earth to explain the observed high-energy positrons. A DM origin implies a cut-off at the dark matter particle mass (or half the mass in the case of decay), while
1.3 Indirect Dark Matter detection

Figure 1.4: The positron fraction measured by AMS-02, PAMELA and FERMI. Data are collected from [14]

if multiple pulsars are the source of the additional positrons, several spectral features would be expected due to the different cut-offs in the spectra of the various pulsars.

Another handle for identifying the origin of the positron excess is the anisotropy in the arrival direction of the positrons or the whole amount of positrons and electrons. If a single nearby source generated the rise in the positron fraction, despite significant loss of directional information due to diffusion, a small anisotropy should be noticed where measuring the angular distribution of the positrons and electrons plus positrons fluxes. Such an anisotropy should appear even in case of multiple sources, dominated by the nearest
and strongest one. The feature that can indicate the most likely candidate should be the direction of the anisotropy, that is towards the Galactic Center for DM (and in any other direction for pulsars).

Currently there are only upper limits on the anisotropy of the cosmic ray \((e^+ + e^-)\) flux from the Fermi LAT \(^{[12]}\) and on the anisotropy of the positron fraction from AMS-02 \(^{[13]}\), but not sufficiently strong to exclude the proposed dark matter or pulsar scenarios.

Another sensitive signature of DM annihilation or decay is provided by the measurement of an excess of anti-protons in cosmic rays. Current measurements of the anti-proton flux yield strong bounds on dark matter models, and have been used to test dark matter interpretations of the Galactic Center gamma-ray excess. The latest results from AMS-02 show the \(\bar{p}/p\) ratio (Fig. 1.5) to be almost constant in the range from 20 GeV up to 450 GeV. This behavior is hard to be explained in terms of secondary production of anti-protons from collisions of ordinary cosmic-rays or pulsar injection as well.
1.3 Indirect Dark Matter detection

Figure 1.5: $\pi/p$ ratio measured by AMS-02 compared with the expected ratio due to secondary production from cosmic-rays.
Chapter 2

The DArk Matter Particle Explorer

2.1 The DAMPE mission

Starting with the remarkable contributions of the Italian-Dutch mission Beppo-SAX (1996-2002) in the study of X-ray Astronomy and its crucial role in resolving the origin of gamma-ray bursts (GRBs), in the last twenty years many important experiments have pushed the observations to higher bands of the energy spectrum. Well known examples are AMS-01 (1998), PAMELA (2006), AGILE (2007), the Fermi LAT (2008), and the more recent AMS-02 experiment on board the International Space Station (ISS), launched in 2011. The Dark Matter Particle Explorer (DAMPE), is a space mission supported by the strategic priority science and technology projects in space science of the Chinese Academy of Sciences (CAS). The satellite was launched on a Long March 2D rocket from the Jiuquan Satellite Launch Center on December 17th 2015 into a sun-synchronous orbit at the altitude of 500 km.

The DAMPE scientific collaboration is composed by institutes from China, Switzerland and Italy. The Chinese institutes involved are the Purple Mountain Observatory (PMO) in Nanjing, the University of Science and Tech-
2.2 The DAMPE Detector

The DAMPE detector (shown in Figure 2.2) consists of 4 sub-detectors, which will be described individually in this section. Looking from top to bottom, the detector is composed of a Plastic Scintillator strip Detector (PSD), a Silicon-Tungsten tracKer-converter (STK), a BGO ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$) imaging calorimeter and a NeUtron Detector (NUD).
2.2.1 The Plastic Scintillator strip Detector (PSD)

The Plastic Scintillator strip Detector has a dual role of serving as anti-coincidence detector and to measure the charge of incident high-energy particles. It can provide precise measurements over a wide range of the electric charge number $Z$, up to $Z=20$.

The PSD has a double layer configuration and is composed in total of 82 modules (Figure 2.3). Each module is equipped with a plastic scintillator bar of 884 mm length with a 28 mm $\times$ 10 mm cross section and the signals are readout by two photomultiplier tubes (PMTs) coupled to the ends of the bar. The modules in a layer are staggered by 8 mm, and the directions of the scintillator bars in the two layers are perpendicular in order to provide respectively alignment in X-Z and Y-Z projection plane. With this dual-view crossed structure, the PSD fully covers an area of 825 mm $\times$ 825 mm and an incident particle will hit at least two modules. This results in a 95% detector efficiency for a single module that leads to an overall efficiency for the whole PSD $\geq 99.75\%$. 
In order to evaluate the PSD measurement capabilities and the choice made for the readout electronics interfaced to the bars, we will consider the relationship between the response of the scintillator to a charged particle and its scintillation efficiency.

When an high energy charged particle passes through a thin plastic scintillator, the interaction with the material produces ionization which generates optical photons, usually in the blue to green wavelength regions [2]. Typical photon yields are about 1 photon per 100 eV of energy deposit. A 1 cm-thick scintillator traversed by a minimum-ionizing particle (MIP) with a release of about 2 MeV/cm will therefore yield \( \approx 2 \times 10^4 \) photons. The resulting photoelectron signal will depend on the collection and transport efficiency of the optical package and the quantum efficiency of the photodetector.

Due to theoretical difficulties in evaluating the quenching effect, the estima-
tion of the light yield in the DAMPE PSD has been based on the direct relativistic heavy ion beam test results of the AMS-02 TOF prototype [15], which also uses the same plastic scintillator material EJ-200 and has the same thickness. The parameters in [15] show that the light output of calcium (Z = 20) is about 270 times larger than that of proton (Z = 1).

In addition, one should consider the broad field view of DAMPE, which accepts a maximum incidence angle of 60°. It indicates that a particle with such an incidence angle has a traversing length and an energy deposit that is double to the perpendicular incident particle, and this extends the maximum light yield to 540 times the light yield of a MIP. Moreover, taking account of the statistical nature of the ionizing process, an estimation of 10% fluctuation (σ) in energy resolution brings to an upper limit of the measurable energy deposition per scintillator bar up to 675 MIPs (required for a 5σ width), and the lower limit of measurement must be smaller than 0.75 MIP.

Taking also into account the noise from photomultiplier tubes (PMT) and electronics, and to ensure the identification of even the smallest signal from noise clearly, the final dynamic range demand for the readout unit of PSD is determined to be 0.1 MIP ~ 675 MIPs. This energy deposition ranges over 4 orders of magnitude in each scintillator bar and cannot be covered by a single readout channel. The system uses a PMT with double dynode outputs coupled to separate the VA160 channels. In this way, two measurement ranges can be achieved with the high-gain dynode channel for small light output measurement and with the low-gain dynode channel for large light output measurement. The scheme of the readout system is shown in Figure 2.4.

Further information about the PSD readout system and calibration tests can be found in [15].
2.2.2 The Silicon-Tungsten Tracker (STK)

The Silicon-Tungsten Tracker (STK) is composed of 6 tracking planes each consisting of two layers of single-sided silicon micro-strip detectors measuring the two orthogonal views perpendicular to the pointing direction of the detector (Fig. 2.5). Three layers of tungsten plates of 1 mm thickness are inserted in front of tracking layers 2, 3 and 4 (the first layer is the closest to the PSD), for enhancing the photon conversion probability. The presence of the tungsten layers is an approach used also in previous space experiments (Fermi [16], Agile [17]). Incoming photons convert into an electron-positron pair in one of the tungsten converters. The use of multiple foils aims to make negligible the multiple scattering effect of the \( e^+e^- \) pair: it can be measured that for high energy particles (> 5 GeV) this effect is negligible if the tungsten thickness is of a few millimeters (\( \theta_0 = 0.08^\circ \) for a 1mm tungsten layer, and 5 GeV particles).

The first layer provides the coordinate of the entry point of the particle in the STK. No additional tungsten foils are placed after layer 4, in order to precisely reconstruct the track of the electron-positron pairs produced by photon conversion in the uppermost layers.

The STK is composed of 192 "ladders", that are modules consisting of 4
2.2 The DAMPE Detector

Figure 2.5: The DAMPE STK scheme. The layer number increases from top to bottom.

single-sided micro-strip detectors (Fig.2.6). The sensor strips are parallel to the longer side of the ladder, and are daisy-chained via micro-wire bonds. Each silicon detector consists of 768 $p^+$-strips implanted in the n-doped bulk. The strips are AC-coupled and biased through polysilicon resistors. The dimensions of the detectors is $95 \times 95 \times 0.32 \text{ mm}^3$.

Figure 2.6: The silicon ladder made of four silicon micro-strip detectors. The front-end electronics is located at one end of the ladder. It reads out 384 strips, with a readout pitch of 242 $\mu$m.

In Table 2.1 are shown some specific information about the silicon detector modules.
The SNR for a MIP crossing a ladder at normal incidence is about 15. The 192 ladders are distributed on seven support trays. Five trays are-equipped on both sides, while the front and the back trays are equipped only on the side facing the interior of the tracker. A layer is composed of 16 ladders, arranged in two rows of 8. A layer is thus composed of an array of $8 \times 8$ detectors. On the double-sided trays, the silicon layer strips are orthogonal to each other.

To sustain the vibrations a light but rigid structure for the support tray has been made of an aluminum honeycomb frame sandwiched between two Carbon Reinforced Polymer (CFRP) face sheets of 0.6 mm thick each (1.0 mm for trays with tungsten). In trays with tungsten, the foil is located immediately above the corresponding tracking layer, to improve the measurement of the conversion point.

The silicon ladders on the bottom surface of a tracker tray and those on the top surface of the tray below form an $X-Y$ tracking plane; the distance between the X and the Y layer of a tracking plane is $\sim 3$ mm. The stack of the 7 trays forms 6 tracking planes that provide 12 measurement points ($6 \times X$ and $6 \times Y$) when a charged track crosses the STK. This structure allows the STK to achieve a spatial resolution better than $50 \mu m$, for incident angles lower than $40^\circ$. 

### Table 2.1: Silicon detector modules information

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip width</td>
<td>$40 \mu m$</td>
</tr>
<tr>
<td>Strip length</td>
<td>93.196 mm</td>
</tr>
<tr>
<td>Strip pitch</td>
<td>$121 \mu m$</td>
</tr>
<tr>
<td>Coupling capacitor</td>
<td>500 pF</td>
</tr>
<tr>
<td>Readout channels</td>
<td>384</td>
</tr>
<tr>
<td>Bias voltage</td>
<td>80 V</td>
</tr>
</tbody>
</table>


More information about the STK readout electronics and performance studies during the assembly phase can be found in [18].

2.2.3 The BGO Calorimeter

The BGO ($Bi_4Ge_3O_{12}$) Calorimeter is composed of 308 BGO crystal bars with dimension of 2.5 cm $\times$ 2.5 cm $\times$ 60 cm each. The crystals are optically isolated from each other and are arranged horizontally in 14 layers of 22 BGO bars. In order to measure the X and Y coordinates information of the particles, each layer is arranged with bars oriented perpendicularly with respect to the direction of the bars in the previous one (Fig. 2.7).

![Figure 2.7: The DAMPE BGO scheme.](image)

In Tab. 2.2 are reported the main properties of the BGO crystal scintillator. The total vertical depth of the DAMPE calorimeter corresponds to about 32 radiation lengths $X_0$, so electrons/positrons and $\gamma$-rays with energies below 200 GeV will deposit almost 100% of their energy inside the
### 2.2 The DAMPE Detector

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$ (g/cm$^3$)</td>
<td>7.13</td>
</tr>
<tr>
<td>$X_0$ (cm)</td>
<td>1.12</td>
</tr>
<tr>
<td>$R_M$ (cm)</td>
<td>2.23</td>
</tr>
<tr>
<td>$dE/dx$ (MeV/cm)</td>
<td>9.0</td>
</tr>
<tr>
<td>$\lambda$ (g/cm$^3$)</td>
<td>22.8</td>
</tr>
<tr>
<td>$\tau_{\text{decay}}$ (ns)</td>
<td>300</td>
</tr>
<tr>
<td>$\lambda_{\text{max}}$ (nm)</td>
<td>480</td>
</tr>
</tbody>
</table>

Table 2.2: Properties of the BGO ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$) inorganic crystal scintillator.

calorimeter, and protons about 40% (see Appendix A).

The scintillation light produced inside the BGO crystal bars is read out by two PMTs located on both ends of the bar. The light asymmetry measured provides a determination of the position of the energy deposition with a resolution of a few cm along the BGO bar.

In order to cover a very large dynamic range of energy deposition, as described previously for PSD, the signals are read out from three different dynodes (Fig. 2.8). Moreover, an attenuation filter is located in front of one of the two PMTs on each bar. In Tab. 2.3 there is a summary of the range covered by different readout channels.

<table>
<thead>
<tr>
<th>Dynode</th>
<th>Energy range S0 end</th>
<th>Energy range S1 end</th>
</tr>
</thead>
<tbody>
<tr>
<td>High gain channel (dy8)</td>
<td>2 MeV - 500 MeV</td>
<td>10 MeV - 2500 MeV</td>
</tr>
<tr>
<td>Medium gain channel (dy5)</td>
<td>80 MeV - 20 GeV</td>
<td>400 MeV - 100 GeV</td>
</tr>
<tr>
<td>Low gain channel (dy2)</td>
<td>3.2 GeV - 800 GeV</td>
<td>16 GeV - 4000 GeV</td>
</tr>
</tbody>
</table>

Table 2.3: Summary of energy ranges covered by readout channels

All these proprieties allow the BGO to perform two main tasks:
2.2 The DAMPE Detector

Figure 2.8: The DAMPE BGO readout scheme.

- to measure the energy deposition of incident particles in a wide energy range;

- to image both the longitudinal and transverse shower development profiles, in order to provide powerful particle identification.

The BGO calorimeter has a leading role in the DAMPE detector. As can be seen in Tab. 2.4, the deep calorimeter length ensures the best performance in terms of energy and angular resolution and allows for a wide range of observation in the energy spectrum, above all the other major space experiments. The sampling of the calorimeter gives also very good capabilities in $e/p$ separation, which will be investigated in this work.

2.2.4 The NeUtron Detector (NUD)

The NeUtron Detector (NUD) is located below the BGO and consists of 4 1 cm-thick Boron-doped plastic scintillator plates of dimensions 19.5 cm $\times$ 19.5 cm, each one readout by a PMT. Neutrons that enter a Boron-doped plastic scintillator undergo the capture process...
2.2 The DAMPE Detector

<table>
<thead>
<tr>
<th></th>
<th>DAMPE</th>
<th>AMS-02</th>
<th>PAMELA</th>
<th>Fermi-LAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e/\gamma$ Energy Res. at 100GeV</td>
<td>1.5%</td>
<td>3%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>$e/\gamma$ Ang. Res. at 100GeV</td>
<td>0.1°</td>
<td>0.3°</td>
<td>1°</td>
<td>0.1°</td>
</tr>
<tr>
<td>Acceptance ($m^2 \cdot sr$)</td>
<td>0.3</td>
<td>0.09</td>
<td>0.002</td>
<td>1.0</td>
</tr>
<tr>
<td>$e/p$ discrimination</td>
<td>$10^5$</td>
<td>$10^6$</td>
<td>$10^4$</td>
<td>$10^3$</td>
</tr>
<tr>
<td>$e/\gamma$ Energy range (GeV)</td>
<td>$5 - 10^4$</td>
<td>$0.1 - 10^3$</td>
<td>0.1-300</td>
<td>0.02-300</td>
</tr>
<tr>
<td>Calorimeter thickness ($X_0$)</td>
<td>32</td>
<td>17</td>
<td>16</td>
<td>8.6</td>
</tr>
</tbody>
</table>

Table 2.4: Comparison of key parameters between DAMPE and other experiments

The process

$$^{10}B + n \rightarrow ^7 Li + \alpha + \gamma \quad (2.1)$$

with a probability inversely proportional to neutron speed, and a time constant for capture inversely proportional to $^{10}B$ loading. Roughly 570 photons in the optical range are produced in each capture [19].

The neutrons produced in hadron-induced showers are thermalized in the BGO calorimeter, in a few microseconds. For the NUD, neutron captures become the dominant source of photon generation beginning 2 μs after the initial shower in the calorimeter.

![Figure 2.9: The DAMPE NUD structure.](image)
Fig 2.9 shows the detailed structure of the NUD, consisting of four boron-loaded plastic scintillators, each equipped with a set of PMTs and their auxiliary circuit. Those plastic scintillators are wrapped with a layer of aluminum film for photon reflection, and anchored in an aluminum alloy framework. Each scintillator is readout by an embedded wavelength shifter directly coupled to a 10-dynode PMT located in the corner: this readout reduces the attenuation in optical transmission and increases photon collection efficiency.
Chapter 3

Beam Test analysis on BGO calorimeter

Two beam tests for the DAMPE engineering qualified model (including all sub-detectors) have been performed in the T9 line of PS and in the H4 line of SPS at CERN (Fig. 3.1). Electrons and hadrons (e.g. proton, pion etc.) were used in the first test (with campaigns on October-November 2014 and June 2015) while in the second were used ions (with a single campaign on March 2015).

Figure 3.2 shows the typical configuration of the charged particle beam test. The external trigger is provided by a set of plastic scintillator detectors (S0, S1 etc.). A set of SSDs (SSD1, SSD2 etc) is used to track the incident particles. The scintillator SV has the function of veto detector in the trigger logic, in order to exclude those events in which the beam particles interacted with matter upstream DAMPE.

Some configurations also included two Cherenkov gas detectors upstream the scintillators as ancillary detectors for particle identification. Finally, to select photons, the magnet is switched on and photons are tagged using the information on the bending angle of their parent electrons.

Since the main purpose of this work is to study the separation of electron
events from proton events inside the BGO calorimeter, a first analysis has been made on electron and proton runs of the beam test campaigns, in order to characterize the calorimeter and verify the expected energy resolution and its capability in resolving the development of electromagnetic showers. The characterization analysis will follow these main steps:

- Parameterization of electromagnetic cascades at different energies
- Energy corrections
- Energy resolution before and after corrections
- Linearity in energy measurement before and after corrections

All the other sub-detectors of the DAMPE satellite will be used as ancillary detectors in event selection and will not be further analyzed.
3.1 Profiles of electromagnetic cascades

Several runs with electron beam have been performed during beam test at different beam momenta. For the parameterization of the cascade profile, the runs with beam momentum of 5 GeV/c and 250 GeV/c will be considered to emphasize the difference between the two longitudinal developments. Fig.3.3 and 3.4 show the distributions of the measured energy deposits in the calorimeter for 5 GeV and 250 GeV electrons, without any correction. The peaks in each spectrum have been fitted with gaussian functions resulting in a total measured energy that is slightly lower than the beam energy.

In describing the electromagnetic shower behavior, the well-known parametrization fully reported in Appendix A has been used, with the scale variables:

\[ t = \frac{x}{X_0} \]  \hspace{1cm} (3.1)

and

\[ y = \frac{E}{E_c} \]  \hspace{1cm} (3.2)
3.1 Profiles of electromagnetic cascades

Energy spectrum at $E_{\text{beam}} = 5$ GeV

Figure 3.3: The BGO energy spectrum for runs at $E_{\text{beam}} = 5$ GeV.

Energy spectrum at $E_{\text{beam}} = 250$ GeV

Figure 3.4: The BGO energy spectrum for runs at $E_{\text{beam}} = 250$ GeV.
where $X_0$ is the radiation length and $E_c$ is the critical energy. The energy deposition profiles are represented in Fig. 3.5 and fitted with the well-known gamma distribution:

$$\frac{1}{E_0} \frac{dE}{dt} = \beta \frac{\alpha^{-1} e^{-\beta t}}{\Gamma(\alpha)}$$  \hspace{1cm} (3.3)

The maximum $t_{max}$ occurs at

$$t_{max} = \frac{(\alpha - 1)}{\beta} = \ln(y) + C_e$$  \hspace{1cm} (3.4)

The fitted parameters result in the following values for $t_{max}$:

$$t_{max}(5\text{GeV}) = 3.52 \quad t_{max}(250\text{GeV}) = 7.80$$  \hspace{1cm} (3.5)

expressed in units of $X_0$.

Figure 3.5: Energy deposition profiles for runs at 5 GeV and 250 GeV

The profiles show that the sampling structure of the BGO can resolve very well the longitudinal development of the electromagnetic shower of an
3.2 Energy correction

As pointed out in sec. 3.1, the energy spectra in Fig. 3.3 and 3.4 show that the energy deposited in the calorimeter is less than the true energy of the incident electrons. This occurs because of possible energy losses in dead zones between adjacent bars of the same layer or between two subsequent layers. The methods adopted for the following corrections are described in [21], [22] and the analysis procedure is from [23].

The S1/S3 method for low energies

Fig. 3.6 shows the two possible incidence configuration for a particle hitting the detector with an angle of 0°.

![Figure 3.6: Possible incidence electron positions.](image)

The first method used to compensate the different energy deposit with
3.2 Energy correction

the hitting position is the S1/S3 method 21 22 that takes into account
the small amounts of dead material between neighboring cells, in particular
when the cells are on the same layer. For the layer \( i \), the maximum energy
deposited in a bar is denoted with \( s_1[i] \), and the sum of \( s_1[i] \) with the energy
deposited in its 2 adjacent bars is denoted with \( s_3[i] \). Then the \( S_1/S_3 \) ratio
is defined as:

\[
\frac{S_1}{S_3} = \frac{\sum_i s_1[i]}{\sum_i s_3[i]} \quad (3.6)
\]

Where the sums run on the BGO layers.

The presence of dead material between the cells reflects on the \( S_1/S_3 \) ratio.
Indeed, as can be noticed in Fig.3.7 for 5 GeV electrons, the \( S_1/S_3 \) ratio is
about 0.5 for particles hitting the calorimeter in the dead space, registering
the highest energy loss, and about 0.9 for particles hitting the bar close to its
center. The relation between the reconstructed energy versus the \( S_1/S_3 \) ratio
can be fitted with a piecewise-defined continuous function \( F(x) \) that consists
of three different linear functions with empirical boundaries:

\[
F(x) = \begin{cases} 
  m_1 x + q_1 & \text{if } a \leq x \leq x_1 \\
  m_2 x + q_1 + x_1(m_1 - m_2) & \text{if } x_1 < x \leq x_2 \\
  m_3 x + q_1 + x_1(m_1 - m_2) + x_2(m_2 - m_3) & \text{if } x_2 < x \leq b
\end{cases}
\]

were \( a \) and \( b \) are the boundaries of the fit range. The energy is corrected
using the fit function \( F(x) \).

This method works fine for low energies and has been applied for runs with
\( E \leq 50\text{GeV} \). For higher energies, the F-Z Method (which will be described
in the next section) resulted to be more effective. In Fig.3.8 the resulted
corrected spectrum has a distribution with a mean value closer to that of the
real beam energy than the raw one.
3.2 Energy correction

Energy reconstructed vs S1/S3

Figure 3.7: Reconstructed energy vs S1/S3 ratio for 5 GeV electrons. The fit with the function $F(x)$ is superimposed to the plot

The F-Z Method for high energies

The F-Z method \[23\] takes into account the lack of energy deposition due to both the gaps between BGO bars of the same layer and the gaps between the layers. The F-Z method uses the following parameters:

- $F$ which is defined as the ratio of energy deposited in the calorimeter against the incident energy;

- $Z_{bary}$ defined as:

$$Z_{bary} = \frac{\sum_i E_i Z_i}{E_{tot}}$$  \hspace{1cm} (3.7)

that is the barycenter of the longitudinal development of the shower, where $E_i$ is the energy deposited in the $i$-th layer and $Z_i$, expressed in radiation lenghts, is the coordinate parallel to the beam.
The F-Z relation indicates that higher $Z_{\text{bary}}$ values correspond to electron going through more BGO crystal layers, which could cause more energy loss in the gaps between the layers.

The F-Z relation for the electron runs at 150 GeV and an incidence angle of $0^\circ$ is shown in Fig. 3.9.

The profile is fitted with a second degree polynomial function to compute the parameters for the energy correction, which were applied resulting in the corrected spectrum shown in Fig. 3.10.

In Table 3.1 are summarized the energy ratios before and after the corrections made with the $S_1/S_3$ method for energies below 50 GeV and F-Z method for higher energies.
3.2 Energy correction

Figure 3.9: F-Z relation for electrons at 150 GeV

Figure 3.10: Raw and corrected spectra for 150 GeV electrons.
### 3.2 Energy correction

<table>
<thead>
<tr>
<th>$E_{\text{beam}}$ (GeV)</th>
<th>Raw (GeV)</th>
<th>Corrected (GeV)</th>
<th>Raw Ratio (%)</th>
<th>Corr. Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.73</td>
<td>2.992</td>
<td>87.4</td>
<td>95.3</td>
</tr>
<tr>
<td>4</td>
<td>3.539</td>
<td>3.91</td>
<td>88.4</td>
<td>97.8</td>
</tr>
<tr>
<td>5</td>
<td>4.64</td>
<td>4.983</td>
<td>92.8</td>
<td>99.66</td>
</tr>
<tr>
<td>10</td>
<td>9.08</td>
<td>9.979</td>
<td>90.8</td>
<td>99.79</td>
</tr>
<tr>
<td>20</td>
<td>17.96</td>
<td>19.969</td>
<td>89.8</td>
<td>99.85</td>
</tr>
<tr>
<td>100</td>
<td>90.076</td>
<td>99.81</td>
<td>90.08</td>
<td>99.81</td>
</tr>
<tr>
<td>150</td>
<td>137.01</td>
<td>149.76</td>
<td>91.9</td>
<td>99.84</td>
</tr>
<tr>
<td>200</td>
<td>181.4</td>
<td>199.54</td>
<td>92.08</td>
<td>99.77</td>
</tr>
<tr>
<td>250</td>
<td>223.03</td>
<td>249.61</td>
<td>91.78</td>
<td>99.84</td>
</tr>
</tbody>
</table>

Table 3.1: Mean values and energy ratios for raw and corrected spectrum fitted with gaussian functions.

### Energy resolution and linearity

In Fig. 3.11 and 3.12 are compared the reconstruction linearity and the energy resolution between raw and corrected data.

It is very clear from fit results that linearity has significantly improved with the corrected values for reconstructed energy.

Improvements are also achieved for the energy resolution, that we can fit with the well known function:

$$\frac{\sigma_E}{E} = \frac{p_0}{\sqrt{E}} + \frac{p_1}{E} + p_2$$  \hspace{1cm} (3.8)

The values of the parameters $p_0$, $p_1$ and $p_2$ obtained from gaussian fits are reported in Tab. 3.2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_0(\text{GeV}^{1/2})$</td>
<td>$(5.18 \pm 0.15) \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$p_1(\text{GeV})$</td>
<td>$(2.17 \pm 0.48) \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$p_2$</td>
<td>$(8.26 \pm 0.61) \cdot 10^{-3}$</td>
</tr>
</tbody>
</table>

Table 3.2: Fitted parameters for energy resolution.
3.2 Energy correction

Figure 3.11: Comparison of reconstruction linearity between raw and corrected data
Figure 3.12: Comparison of energy resolutions between raw and corrected data
Chapter 4

The $e^-/p$ separation multivariate analysis on DAMPE BGO calorimeter

The particle identification in High Energy Physics (HEP) experiments is one of the most interesting challenges that pushes to the limit the detectors performance and the data-analysis techniques. One of the main interesting research topics in space-based experiments is about the electronic component of Cosmic Rays (CR), which carries important physics information about the origin and propagation of CR and is complementary to the hadronic component. Due to their strong energy losses, electrons and positrons at high energies are unique probes to study the CR source property in the galactic neighborhood.

Charged cosmic rays between 1 GeV and 1 TeV observed at Earth, are made substantially of protons ($\sim 90\%$), Helium ($\sim 8\%$) and heavy nuclei ($\sim 1\%$). Electrons and positrons constitute respectively $\sim 1\%$ and $\sim 0.1\%$ of the total CR flux. The main challenge in the measurement of the electronic component is the natural low signal/background ratio ($e^-/p \sim 10^{-3} - 10^{-2}$ and $e^+/p \sim 10^{-4} - 10^{-3}$).

Along with the pursue in developing high precision detector measurements,
new data-analysis techniques are being widely used in particle physics such as multivariate analysis with machine learning. In the HEP community, that mostly adopts the ROOT Data Analysis Framework [24], there is also a wide use of a ROOT-based library for machine learning named TMVA [25] (Toolkit for Multivariate Data Analysis). TMVA is a standalone project that provides a ROOT-integrated machine learning environment with multivariate classification techniques that range from decision trees to simple neural networks.

After a first initial use of the TMVA, due to the limited choice of available algorithms, it has been chosen for this work to develop the analysis using python libraries that include two state-of-the-art tools and algorithms widely used in the machine learning community: Scikit-learn [26] and Google TensorFlow TM [27].

Scikit-learn is a set of Python-based libraries which combines a user-friendly interface with a highly optimized implementation of a large variety of learning algorithms in addition to many convenient functions to pre-process data and to fine-tune and evaluate the models built.

TensorFlow is an open source (since late 2015) software library in C++ and Python for numerical computation using data flow graphs. It was originally developed by researchers and engineers working on the Google Brain Team within Google’s Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

In this chapter an approach to the $e^-/p$ separation analysis using machine learning techniques will be described, developing a customized classification algorithm based on well known learning algorithms used in computer vision
and data science.

### 4.1 The imaging calorimeter approach with Beam Test data

The first stage of the analysis has involved the use of data from the beam test campaign described and analyzed in the previous chapter. The amount of BT data is not enough to accumulate large enough statistics to properly train and test machine learning algorithms and obtain results that truly show the real discrimination power of signal/background rejection of the detector, so this first analysis has the purpose of highlighting the sampling proprieties of the DAMPE BGO calorimeter in order to undertake the pattern recognition of the particle shower inside the detector.

The idea behind the shape-recognition of the particle shower comes from the good sampling capability of the DAMPE calorimeter with its 308 bars and the double-view arrangement to create projections in X-Z and Y-Z plane. Fig. 4.1 and 4.2 show respectively the display of an electron and a proton event with a 150 GeV energy deposit in the BGO. The X-axis of the two figures corresponds to the horizontal positions (either along X or Y) of the BGO crystals in a layer. The Y axis of the two figures corresponds to the number of the BGO layers. A comparison of the two figures shows that the longitudinal and transverse profile of the shower are clearly different.

To verify this potentially powerful classification, a *supervised* machine learning algorithm has been developed using Scikit-learn libraries. The following sections will provide either a general description of the algorithm structure and it’s application on a specific BT campaign data-set.
4.1 The imaging calorimeter approach with Beam Test data

Figure 4.1: Composite view of a 150 GeV electron shower in DAMPE BGO

Figure 4.2: Composite view of a 150 GeV proton shower in DAMPE BGO

4.1.1 Data selection from beam test campaign

The choice of the beam test campaigns to apply this classification analysis was mainly driven by the amount of data available at each beam energy. Regarding electrons, the largest statistics has been collected in runs with primary particles of 150 GeV/c momentum. In this campaign from the October 2014 BT, there are approximately $2 \times 10^4$ electron events recorded with incident angles of $0^\circ$ and $30^\circ$. Pre-selection cuts have been made looking for events with only one reconstructed track in
4.1 The imaging calorimeter approach with Beam Test data

the STK and a deposited energy in the BGO greater than 120 GeV, in order to avoid electrons that started showering before entering the BGO.

Figure 4.3: Green area: selected electron events with one reconstructed track in STK and BGO energy greater 120 GeV.

As one can notice in Figure 4.3 discarded events consist of particles with much lower energy deposition, i.e. particles pre-showering in the STK or beam contaminants. After the selection, about $1.1 \times 10^4$ events survived.

Since the most important information about a triggered event in DAMPE comes from the energy deposit in BGO, the goal is to distinguish an electron from a proton with the same energy release in BGO. Following this concept and the need of collect an amount of proton events comparable with that of electrons, data from June 2015 Beam Test campaign have been selected. With the same condition of a unique reconstructed track in STK, selection cuts requesting a deposited energy in BGO between 120 GeV and 150 GeV
over the $\sim 6 \cdot 10^5$ events with primary protons at 400 GeV/c momentum and incident angles of $0^\circ$ and $15^\circ$ result in a total of $\sim 6 \times 10^4$ (Figure 4.4).

Figure 4.4: Selection 400 GeV primary protons events with energy deposit in BGO between 120 GeV and 150 GeV.

The selected electron and proton events, represent respectively the *Signal* and the *Background* for the binary classification, which will be described in the next sections.

### 4.1.2 Supervised dimensionality reduction via Linear Discriminant Analysis

The set of variables that are used to feed the algorithm are the values of energy deposits in individual BGO bars. Therefore a total of 308 starting variables are used as input for each event (154 if we consider single X-Z or Y-Z view). For what concerns this application on BT data there will be
no difference in considering a single view or the composite one, because the
beam tests measurements are carried out at fixed incidence angles, with most
of the events at $0^\circ$ incidence angle and impact point at $Y = 0$.

Due to the huge number of input variables and to the fact that for each event
the shower energy is deposited in a limited region of interest, the first step
is to perform a dimensionality reduction, with the goal of keeping most of
the relevant information. In machine learning application there are two main
ways to perform this task: the *Principal Component Analysis* (PCA), used
for unsupervised models, and the *Linear Discriminant Analysis* (LDA) for
supervised ones.

Since the data used are from BT campaign, after selection cuts on ancillary
detectors, a specific event inside the BGO can be precisely pre-labeled, and
the choice is to use a supervised model. It means that in the training phase
the model parameters will be fitted taking in account also the event label,
trying to maximize the differences between the two classes of events, that we
indicate as *Signal* ($e^-$) and *Background* ($p$).

The goal in LDA is indeed to find the feature subspace that optimizes class
separability, and the key steps can be summarized as follows:

1. Standardize the $d$ -dimensional dataset ($d$ is the number of features)
2. For each class, compute the $d$ -dimensional mean vector.
3. Construct the between-class scatter matrix $S_B$ and the within-class
scatter matrix $S_W$.
4. Compute the eigenvectors and corresponding eigenvalues of the matrix
$S_W^{-1}S_B$.
5. Choose the $k$ eigenvectors that correspond to the $k$ largest eigenval-
ues to construct a $d \times k$ -dimensional transformation matrix $W$; the
4.1 The imaging calorimeter approach with Beam Test data 55

... eigenvectors are the columns of this matrix.

6. Project the samples onto the new feature subspace using the transformation matrix $W$.

To evaluate the number of linear discriminants to extract, solutions of the eigenvalues problem have been evaluated. The discriminatory power of each class is carried by the linear discriminants (eigenvectors), and therefore only the non-zero eigenvalues are extracted.

In this case, only two eigenvalues are found to be non-zero, and the distribution of the LDA components for each class events are showed in Fig 4.5.

![Figure 4.5: LDA components for signal and background.](image-url)
4.1 The imaging calorimeter approach with Beam Test data

The performed LDA analysis resulted in a dimensionality reduction from 308 features to 2, with very good separation power between the two classes (similar results can be found using only a single view of the calorimeter). Despite the lower contribution from the second component (vertical axes) of the LDA, the behavior of the distributions on that dimension can’t be ignored in the search of an optimal cut to separate the two classes. That’s why a more sophisticated estimator is required, and this is the most suitable situation for Support Vector Machines.

4.1.3 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are widely used algorithms in classification problems. Their approach belong the branch of the so called "Kernel methods" because the goal is to find the hyperplane that effectively divides the class representation of data. Hyperplanes can be defined as a generalization of a line in 2-Dimensions and a plane in 3-Dimensions. The optimization goal consists in maximizing the margin, that is defined as the distance between the separating hyperplane (decision boundary) and the training samples that are closest to this hyperplane, which are the so-called support vectors (Fig. 4.6).

The kernels available for SVM does are not restricted to the linear case. In the conventional SVM implementations and in particular for Scikit-learn libraries, there are the following kernel types:

- **linear**: $\langle x, x' \rangle$
- **polynomial**: $(\gamma \langle x, x' \rangle + r)^d$
- **rbf**: $\exp(-\gamma \|x - x'\|^2)$
- **sigmoid**: $\tanh(\gamma \langle x, x' \rangle + r)$
The choice of the best kernel and the tuning of the hyperparameters has been made via grid search. The implementation of a grid search consists in a brute-force exhaustive search paradigm where one specifies a list of values for different hyperparameters, and the computer evaluates the model performance for each combination of those to obtain the optimal set.

The hyperparameters to be tuned for an SVM classifier are essentially the $\gamma$ coefficient and the inverse regularization parameter $C$ (known also as penalty parameter).

The $\gamma$ parameter can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. The $C$ parameter trades off misclassification of training examples against simplicity of the decision surface. A low $C$ makes the decision surface smooth, while a high $C$ aims at classifying all training examples correctly by giving the model freedom to select more samples as support vectors.

Since the grid search selected a linear SVM kernel as best model for the data, the only tuned parameter has been the penalty parameter $C$, at which
is assigned the best-fitted value of 0.1.

Once defined the best kernel with tuned hyperparameters, the model has been trained on the training samples and then tested on the remaining test samples.

**Overtraining check**

One of the most common mistakes in developing a machine learning algorithm is to fit the model in a way perfectly suited for the training samples resulting in a loss of generality and a consequently decrease of the model performance on test data. For this reason it is very important to check carefully the absence of this kind of bias, comparing for example the distributions of scores given at the training data during the training phase and that of the scores of the test data during the test phase (Fig. 4.7).

A useful tool to quantify the distance between the empirical distribution functions of two samples (train and test) is the Kolmogorov-Smirnoff test. This is a two-sided test for the null hypothesis that 2 independent samples are drawn from the same continuous distribution. Applying the K-S test to the distributions in Fig. 4.7 the following p-values result for the two classes:

- $p_{KS}(\text{Signal}) = 0.442$
- $p_{KS}(\text{Background}) = 0.619$

The high p-values obtained allow to keep the null-hypothesis that for each class, the samples belong to the same distribution.
4.1 The imaging calorimeter approach with Beam Test data

Figure 4.7: Distribution of output scores of train and test sample using the tuned Support Vector Classifier

ROC Curve

The Receiver operator characteristic (ROC) graph is the first useful tool to evaluate the model based on its performance with respect to the false positive and the true positive rates, which are computed by shifting the decision threshold of the classifier. The diagonal of a ROC graph can be interpreted as random guessing, and classification models that fall below the diagonal are considered as worse than random guessing. A perfect classifier would fall into the top-left corner of the graph with a true positive rate of 1 and a false positive rate of 0.

Based on the ROC curve, one can then compute the so-called area under the
4.1 The imaging calorimeter approach with Beam Test data

curve (AUC) to characterize the performance of a classification model\cite{28}. The ROC curve for this classifier’s outputs is shown in Fig. 4.8.

![ROC Curve](image)

**Figure 4.8:** ROC Curve for the SVC classifier.

The computed AUC is 0.9991 and the following contamination rates have been calculated at different thresholds and reported in Tab.

<table>
<thead>
<tr>
<th>Signal Efficiency</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>95%</td>
<td>$\leq 10^{-4}$</td>
</tr>
</tbody>
</table>

Where the Signal Efficiency is defined as:

$$SE = \frac{TP}{FN + TP} \quad (4.1)$$
4.2 Deep Learning classification approach with on-orbit data

and TP and FN are the numbers of True Positive and False Negative events. This encouraging results confirmed the great capabilities of particle identification with the BGO calorimeter and lay the foundations for a more sophisticated analysis with application on real on-orbit data that will be described in the following section.

4.2 Deep Learning classification approach with on-orbit data

The $e^-/p$ separation analysis on the events observed on-orbit by DAMPE requires a more complex approach because of the increasing randomness of events geometry with respect to the Beam Test case. The analysis developed consists in five major parts:

- **Good MC events selection**: selection of Monte Carlo events usable to train the algorithm for particle identification.

- **Random Forest Classifier**: developing of decision-trees-based classifier with the Random Forest techniques in machine learning.

- **Convolutional Neural Network Classifier**: developing of a classifier built with the Convolutional Neural Network structures of deep learning.

- **The meta-classifier**: The two developed classifiers are then used together as a single meta-classifier that combines the decisions of each one, enhancing the decision accuracy.

- **Test on on-orbit flight data**: The meta-classifier is finally tested on events observed in space.
4.2 Deep Learning classification approach with on-orbit data

4.2.1 Monte Carlo events generation and pre-selection

The analysis of BT data on particle identification has a non-negligible limitation due to the lack of statistics and to the fixed directions of the particles entering the detector. To ensure a good classification capability of the on-orbit observed events by DAMPE, a more general training sample must be used. Since the observed data cannot be used to train a supervised algorithm, because of the unknown nature of the events, a Monte Carlo data-set is necessary to train the classifier.

The DAMPE collaboration is currently providing a complete MC production that includes simulated on-orbit events with protons, electrons, gamma rays and some ions as primary particles. In order to increase the statistics and compare the results with the previous BT $e/p$ analysis, as part of the apprenticeship an additional huge amount of data has been produced with the following configurations:

- **Primary electrons** with power-law energy distribution between 150 GeV and 155 GeV (1.438 · 10$^7$ events);

- **Primary protons** with power-law energy distribution between 400 GeV and 450 GeV (3.26 · 10$^7$ events).

A preliminary selection among the events has been made in order to consider only events with a clear shower development in the calorimeter, removing all those with which the incoming direction with respect to the orientation of the detector, have a non negligible part of the shower outside the BGO or they didn’t pass through the STK.

This selection of "clean" events, is ruled by cuts on the following variables:
4.2 Deep Learning classification approach with on-orbit data

BGO deposited energy

The first variable for selection is the energy deposition inside the calorimeter, and the selection cuts are:

- $130\,\text{GeV} \leq E_{\text{BGO}} \leq 150\,\text{GeV}$ for **electrons**
- $100\,\text{GeV} \leq E_{\text{BGO}} \leq 180\,\text{GeV}$ for **protons**

Angular separation of reconstructed tracks

In order to verify the effective passage of the particle through both STK and BGO, we define for each view (XZ and YZ) $\theta_{\text{STK}}$ as the angle of the best reconstructed track in the silicon tracker and $\theta_{\text{BGO}}$ as the angle of the reconstructed track in the calorimeter. The selection cut is imposed with the geometrical condition:

$$|\Delta \theta_i| = |\theta_{\text{STK},i} - \theta_{\text{BGO},i}| \leq 10^\circ \quad (4.2)$$

where $i$ stands for the individual XZ or YZ view. The condition on $\Delta \theta$ must be satisfied on both views.

Transverse Containment variable

The last variable used for the pre-selection of clean events is defined as:

$$\frac{\langle E^i_{\text{BGO}} \rangle}{E^i_{\text{layer, max}}} \quad (4.3)$$

where $\langle E^i_{\text{BGO}} \rangle$ is the mean deposited energy per BGO layer, $E^i_{\text{layer, max}}$ is the energy deposited in the layer with the maximum deposition in the and the index $i$ is indicates the XZ or YZ view.

The meaning of this variable is to quantify the amount of electromagnetic cascade inside the BGO for each event: the lower the value, the more the
Figure 4.9: Distribution of the transverse containment variable for MC electrons, the red dashed line shows the cut value applied. Events with a value $\geq 0.22$ are selected.

The distribution of this variable for electrons is shown in Fig. 4.9.

The red dashed line shows the cut value applied to the selection of electrons in the XZ view. Events with a value lower than 0.22 have been rejected and similar values are used for both views of protons and for the YZ view of electrons.

In Figures 4.10 and 4.11 are shown the spectra of the selected events respectively for electrons and protons, and in Table 4.1 the efficiencies of these cuts are summarized.

<table>
<thead>
<tr>
<th>Particle</th>
<th>Simulated events</th>
<th>Selected events</th>
<th>Cut efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^-$</td>
<td>$1.438 \cdot 10^7$</td>
<td>167821</td>
<td>1.2%</td>
</tr>
<tr>
<td>$p$</td>
<td>$3.26 \cdot 10^7$</td>
<td>108490</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Table 4.1: Cut efficiencies for events pre-selection.
4.2 Deep Learning classification approach with on-orbit data

Figure 4.10: Selected spectrum (in red) for MC electrons on the overall spectrum of the MC production (in blue)

From now on, the sample size used for the analysis will be of $10^5$ events for each particle; the 60% of the sample will be used for training and the remaining 40% for test.

4.2.2 The Random Forest Classifier

The Random Forest (RF) technique is based on the idea of ensamble learning that combines weak learners to build a more robust model, a strong learner that has a lower generalization error, and is less liable to overtraining. Random Forests are based on the well known Decision Trees, which are learning methods that predict the value of a target variable by learning simple decision rules inferred from the data features. The use of an ensamble of Decision Trees and the adoption of techniques based on randomness constitute the core of Random Forest algorithms which structure can be summarized in
4.2 Deep Learning classification approach with on-orbit data

four simple steps\cite{28}:

1. Draw a random bootstrap sample of size $n$ (randomly choose $n$ samples from the training set with replacement).

2. Grow a Decision Tree from the selected bootstrap sample and, at each node:
   - Randomly select $d$ features without replacement.
   - Split the node using the feature that provides the best split according to the objective function.

3. Repeat the previous steps 1 and 2, $k$ times.

4. Assign the class label among all decisions by majority vote.
4.2 Deep Learning classification approach with on-orbit data

The property that mostly characterizes the RF is indeed in point 2 where, unlike the single Decision Tree, instead of evaluating all features to determine the best split at each node, only a random subset of those is considered. The features selected for this analysis are a set of parameters that can be extracted from the reconstruction of each event inside the BGO calorimeter. At each training session, one can evaluate the discrimination power of each feature measuring the contribution to the Random Forest as the averaged impurity decrease computed from all decision trees in the forest.

In the following 9 selected features ranked by their relative contribution to the classification are reported.

- **F-Value at layer 14** defined as:
  \[
  F - Value(14) = \frac{RMS^2(14) \times E_{BGO}(14)}{E_{TOT}^{BGO}}
  \]  
  (4.4)

- **Energy in layer 13 and 14** defined as:
  \[
  E_{13,14}^{BGO} = E_{13}^{BGO} + E_{14}^{BGO}
  \]  
  (4.5)

- **RMS^2 at maximum of the shower**, with \( RMS^2 \) defined as:
  \[
  RMS^2[i] = \sum_{j=0}^{21} E_{BGO}[i][j] \times (BGOHitPos[i][j] - CoG)^2
  \]  
  \[
  \sum_{j=0}^{21} E_{BGO}[i][j]
  \]  
  (4.6)

In the previous formula \( BGOHitPosition \) is an X-coordinate if the layer is along X, or Y-coordinate if is along Y. \( E_{BGO} \), a 2-D array of size [14][22], which indicates the energy deposited in the BGO bars, where 14 is the number of layers and 22 the number of bars for each layer. \( CoG \) stands for "Center of Gravity", which is defined as

\[
CoG[i] = \sum_{j=MaxEnBar-1}^{MaxEnBar+1} (E_{BGO}[i][j] \times BGOHitPos[i][j])
\]  
\[
\sum_{j=0}^{21} E_{BGO}[i][j]
\]  
(4.7)
4.2 Deep Learning classification approach with on-orbit data

- **Total number of hits in the BGO**: the number of bars with deposited energy in each event;

- **Energy in the first 4 layers** defined as:
  \[
  E_{0,3}^{BGO} = \sum_{i=0}^{3} E_i^{BGO}
  \]  
  \[\text{(4.8)}\]

- **Energy core 3 vs energy core 5 ratio**: is the ratio between the two quantities that measure the energy deposited in a cylinder of one Molière radius (Energy core 3) or two Molière radii (Energy core 5);

- **Energy ratio of the first 4 layers** defined as:
  \[
  E_{r0,3}^{BGO} = \frac{\sum_{i=0}^{3} E_i^{BGO}}{E_{TOT}^{BGO}}
  \]  
  \[\text{(4.9)}\]

- **Energy in layer 0**: is the energy in the first layer;

- **Reduced $\chi^2$**: the reduced $\chi^2$ of the reconstructed track in the BGO;

In Figures 4.12, 4.13 and 4.14 the distributions for the features that give the major contribution to the overall decision of the forest are shown. The relative contributions for all the features are summarized in Table 4.2.

Despite the low value of their relative contribution, the last features have an essential role that can be quantified if they are excluded from the classifier, resulting in a significant decrease of the classifier’s performance.

**Training and test of the Random Forest classifier**

The classifier built with this well set of variables has been trained with 1000 estimators (trees in the forest) using as objective function at each decision node of the single trees the well known *Gini index* (or Gini coefficient)
4.2 Deep Learning classification approach with on-orbit data

<table>
<thead>
<tr>
<th>Feature</th>
<th>Relative contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Value at layer 14</td>
<td>35.6 %</td>
</tr>
<tr>
<td>Energy in layer 13 and 14</td>
<td>22.5 %</td>
</tr>
<tr>
<td>RMS$^2$ at maximum of the shower</td>
<td>17.0 %</td>
</tr>
<tr>
<td>Total number of hits in the BGO</td>
<td>8.1 %</td>
</tr>
<tr>
<td>Energy in the first 4 layers</td>
<td>5.76 %</td>
</tr>
<tr>
<td>Energy core 3 vs energy core 5</td>
<td>4.5 %</td>
</tr>
<tr>
<td>Energy ratio of the first 4 layers</td>
<td>3.2 %</td>
</tr>
<tr>
<td>Energy in layer 0</td>
<td>2.1 %</td>
</tr>
<tr>
<td>Reduced $\chi^2$</td>
<td>1.1 %</td>
</tr>
</tbody>
</table>

Table 4.2: Features relative contribution to overall decision in the Random Forest classifier.

![Graph showing distribution for the F-value feature at layer 14 for electrons (red) and protons (blue).](image)

Figure 4.12: Distribution for the F-value feature at layer 14 for electrons (red) and protons (blue)

that can be expressed at each node, in terms of the node class populations $N_e, N_p$ and the total node population $N$:

$$Q_{Gini} = \frac{4}{N} \sigma_{binomial}^2 = 4 \frac{N_e N_p}{N^2} = 4 \frac{N_e (N - N_e)}{N^2} \in [0, 1]$$

(4.10)
4.2 Deep Learning classification approach with on-orbit data

Figure 4.13: Distribution for the $E_{13,14}^{BGO}$ feature for electrons (green) and protons (blue)

$Q_{Gini}$ of a node is zero for the ideal case that only one class is present in the node ($N_e = 0$ or $N_p = 0$). The Gini index of the split is calculated by adding the Gini indices of the two successor nodes (denoted by left and right node) and scaling the result to $[0,1]$:

$$Q_{Gini} = 2\left(\frac{N_{e,\text{left}} N_{p,\text{left}}}{N_{\text{left}}} + \frac{N_{e,\text{right}} N_{p,\text{right}}}{N_{\text{right}}}\right)$$  \hspace{1cm} (4.11)

In order to minimize the variance of the population of electron and protons one has to choose the smallest $Q_{Gini}$ that will naturally purify the sample. At each node has been considered a subset of 3 features ($d = \sqrt{n} = 3$) and the minimization of the Gini index provides both the choice of the feature and the split value to be used\textsuperscript{[30]}.

After the training with the 120000 training events (60000 for each particle), the remaining 80000 events have been tested and the resulting ROC curve is shown in Fig.4.15 with an AUC very close to unit.
4.2 Deep Learning classification approach with on-orbit data

![Distribution for the $RMS^2_{max}$ feature for electrons (green) and protons (blue)](image)

Figure 4.14: Distribution for the $RMS^2_{max}$ feature for electrons (green) and protons (blue)

**Overtraining check**

To carefully check if the model is affected from overtraining, a wide-used method for Random Forest classifiers is to look at the *learning curves* that consist in plotting the model training and validation accuracies as functions of the training set size\(^{[28]}\). This kind of approach is very useful with models like RF with too many degrees of freedom and parameters.

To create these learning curve, the chosen validation method has been the *K-fold cross-validation*. In K-fold cross-validation the training dataset is randomly split into $k$ folds without replacement, where $k-1$ folds are used for the model training and one fold is used for testing. This procedure is repeated $k$ times so that $k$ models and their performance estimates are obtained. After that, the average performance of the models is calculated to obtain a performance estimate that is less sensitive to the subpartitioning of the training
4.2 Deep Learning classification approach with on-orbit data

Figure 4.15: ROC curve for the Random Forest Classifier on MC data

Although this validation technique is very expensive in terms of computational resources and time, it is a powerful way to verify the solidity of the model and the stability of its performance.

With a $k$-value $= 10$, the resulting learning curves are shown in Fig. 4.16. The model reaches a false positive rate of $6 \times 10^{-4}$ at 90\% signal efficiency. This is yet a good classification performance with low contamination in the output signal sample, but can be further improved with the implementation of a classifier for the pattern recognition on the imaging calorimeter.

4.2.3 The Convolutional Neural Network (ConvNET)

As shown at the beginning of this chapter, the DAMPE BGO calorimeter exhibits good imaging properties and provides for a good resolution of the
developing of the electromagnetic cascade of a showering particle.

With such a good resolution of the electromagnetic cascade, an event can in principle be also classified looking at the shower shape, which (as is well known and also reported in Appendix A) has a slightly different shape if the primary particle is an electron or a proton.

Embedding this concept in a typical analysis of particle identification, a machine learning approach is essential and, in particular, the approach should belong to that branch that involves Neural Networks and has a wide range of application in computer vision: deep learning.

State-of-the-art computer vision mostly adopts for image and pattern recognition the powerful architecture of Convolutional Neural Networks (ConvNets).
A ConvNet for the DAMPE BGO calorimeter

The ConvNet for the DAMPE BGO calorimeter has been built during this work using the powerful *Google-Tensorflow*\(^{[27]}\) libraries, already mentioned at the beginning of this chapter. The developed network scheme of the principal macro-layers is illustrated in Fig. 4.17 and follows the architecture described in \(^{[31]}\), with some changes to adapt at the specific case of the DAMPE imaging calorimeter.

![Diagram](image)

**Figure 4.17:** Scheme of the ConvNET built for DAMPE BGO imaging calorimeter.

To better understand the network architecture and its operations, the most critical steps of the implemented ConvNET will be now illustrated.
The input image

Unlike regular neural networks, ConvNETs cannot have as input stage a simple vector of features to fully describe an image. Their input is assumed as a three-dimensional vector with dimensions called width, height and depth. In this particular case, the image has the dimension of a single-view projection of the calorimeter, so the input size of the first two dimensions will be (7,22), corresponding to the layers and the bars forming the X-Z and Y-Z view of the event. The third dimension is referred to the information contained inside each pixel, and can be for example 1 if the network is fed with a greyscale image, or 3 for a RGB one. In this case the pixel value is the value of deposited energy inside each bar, so it can be assumed as a single channel of information and therefore the input size will be (7,22,1).

Convolutional layer

The Convolutional layer constitutes the core of a ConvNET and its main operations are summarized in the following steps[32] represented in Fig.4.18.

- A kernel (shaded area) which slides across the input feature map (light blue).

- At each location, the product between each element of the kernel and the input element overlapped with the kernel is computed and the results are summed up in a linear combination to obtain the value of the output feature map in the current location.

- The procedure can be repeated using different kernels to form as many output feature maps as desired.

In the first convolutional layer of this network, 16 kernel-filters are applied to the input image of size (7,22,1). For each position of each filter over the
4.2 Deep Learning classification approach with on-orbit data

input image, the dot-product is being calculated between the filter and the image pixels, which results in a single pixel in the output image.

This convolution of a filter over the image results in a new image being generated. In the procedure summarized in Fig 4.18 two important elements can be noticed. The first one is the step size when moving the filter across the image, that is usually called "stride" and in our network has been set to 1. The second one is a characteristic frame of zero-value pixels around the image (in white), called "zero-padding", that allows the filter to cover enough positions to obtain an output image of the same dimension as the input one. After the convolution of all filters, the output for each image is a 4-D tensor of size (7,22,1,16) that contains a stack of the results for the convolution of each filter over the image.

In the second convolutional layer, 36 filters are applied to the input volume coming from the pooling layer.

**Pooling layer**

When a ConvNET is formed by multiple convolutional layers, it is convenient to have a pooling layer in between the convolutional layers. The pooling layer is responsible for reducing the chances of over-fitting by reducing the spatial size of the input volume.

The pooling operation is similar to convolution, but instead to compute the
linear combination of all the dot-products of filter weights and pixel values, it evaluates either the maximum value or the average of all values depending on whether "max pooling" or "average pooling" is used.

In this case a $2 \times 2$ max pooling has been implemented between the convolutional layers in order to downsample the width and height of the 4-D volume to a size of (4, 11). This procedure yields two major benefits:

1. The elimination of non-maximal values reduces computations for the following layers. For a $2 \times 2$ max pooling about 75\% of parameters are removed.

2. Since it combines a dimensionality reduction with the highlighting of the pixels with higher values, the pooling reduces the pattern always to a region of interest, providing a basic form of translation invariance in the output feature maps.

**Fully connected layer**

The fully connected layer works as an interface between the 4-D tensors produced in the convolutional layers and the output score that the network must provide. The preliminary stage for a fully connected layer is to flatten the 4-D tensor in a 2-D vector. After the flattening operation, the fully connected layer computes a product with a weight matrix and provides a 2-elements output corresponding to the number of classes and applies an activation function to obtain an output score for each class. These outputs are the so called "logits" and can be considered as a raw output from the network.
4.2 Deep Learning classification approach with on-orbit data

Softmax layer

The scores provided by the fully connected layer allow to estimate how likely an image belongs to each of the two classes, but they are often very difficult to interpret, and need to be normalized. The normalization procedure is made using the so-called "Softmax function" defined as (for J classes):

\[ S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{J} e^{y_j}} \]  

(4.12)

where \( i \) ranges from 1 to \( J \) and \( y_i \) is the score of the \( i \)-th class.

The Softmax function assigns a proper probability in the interval \([0,1]\) to each class. It is straightforward to apply a threshold at each probability and estimate the classification power of the network.

Network optimization

As in any machine learning problem, all the hyperparameters have to be tuned in the training phase and must follow an optimization criterion. One of the most peculiar mechanism in neural networks is their continuous optimization of weights and biases at each iteration. In this network the cross-entropy has been used as optimization function.

Indicating with \( S \) the vector of class probabilities from the Softmax layer and with \( L \) the vector with the true class label for the event (one-hot-encoded vector with all elements set to zero except the element corresponding to the true class set to 1) the cross-entropy function is defined as:

\[ D(S, L) = - \sum_i L_i \log(S_i) \]  

(4.13)

and corresponds to the "distance" between the predicted labels and the true ones. The cross-entropy is an always positive continuous function, which goes to zero if the predicted output of the model exactly matches the desired
output. The goal of optimization is therefore to minimize the cross-entropy by changing the variables of the network layers.

In the optimization process, the cross-entropy function is estimated for every image and the average cross-entropy over all the classified images is computed and minimized at each iteration. The minimization algorithm used is the ADAM algorithm, a gradient-based optimization method for stochastic objective functions.

**ConvNET overtraining check and performance**

The ConvNET has been trained with the same events used for the Random Forest. Two networks have been trained respectively on the X-Z view and Y-Z view images for each event, and the final scores for each event are computed as the average of the individual scores.

To check for overtraining, it is useful to look at the "accuracy" for the training and test samples at each batch (group of images analyzed at the same time) iteration (Fig.4.19), which is defined as:

\[
ACC = \frac{TP + TN}{FP + FN + TP + TN}.
\]  

with FP, FN, TP TN the number of False Positive, False Negative, True Positive and True Negative.

It is easy to notice that the accuracy reaches very fast values very close to 1 for training and test samples, and the difference between the accuracies for the two samples are not significant, thus ensuring that the model is not affected by overfitting on the train sample. In fact, in this case one should notice a significant reduction in accuracy on the test sample.

The resulting ROC curve is shown in Fig.4.20.
4.2 Deep Learning classification approach with on-orbit data

Figure 4.19: Accuracy for train and test samples.

The model reaches a false positive rate of $7 \times 10^{-4}$ at 90% signal efficiency, comparable to that reached with the Random Forest classifier.

The next step is to form a *meta-classifier* formed by these two models that will have overall performance better than the two individual models.

4.2.4 A meta-classifier with Random Forest and ConvNET

The meta-classifier will use the two models implemented in the same algorithm following this procedure:

- Two ConvNETs are trained each one on the single-view image of each event;
- The Random Forest is trained independently on the selected features;
4.2 Deep Learning classification approach with on-orbit data

Figure 4.20: ROC curve for the Convolutional Neural Network on MC data

- The first classification score is computed with a majority vote between the two ConvNETs;

- The output score from the two ConvNETs and the output score of the Random Forest produce the overall score by taking the average score for each class from the two classifiers.

In Fig. 4.21 and 4.22 are shown the resulting ROC curve and the scores distributions of the meta-classifier.

The model based on the meta-classifier has better performance than both individual classifiers, reaching a false-positive contamination rate of only $7.5 \times 10^{-5}$ at 90\% signal efficiency and of $2.7 \times 10^{-4}$ at 95\% signal efficiency, almost an order of magnitude less than the rates for individual classifiers.
4.2 Deep Learning classification approach with on-orbit data

4.2.5 First application on real observed data

As final step of this analysis, a first preliminary test on real data observed by DAMPE has been made looking at a short time-window of 45 days of data between August 10th and September 24th of 2016. This test must be considered as a simple discrimination between events producing an electromagnetic cascade and events producing hadronic shower inside the BGO calorimeter, because no proper cuts have been made on data to reject contamination from photons, ions and other particles.

To perform the same geometric and spectral pre-selection made on Monte Carlo data, the following cuts have been used:

- Energy deposited in BGO between 120 GeV and 180 GeV (Fig.4.23);
4.2 Deep Learning classification approach with on-orbit data

Figure 4.22: Distributions of the meta-classifier score for Signal and Background events.

- Difference in STK and BGO tracks angles less than 10° on both views:
  \[ |\Delta \theta_i| = |\theta_{STK,i} - \theta_{BGO,i}| \leq 10^\circ \]  
  (4.15)

- Cuts on the containment variables (Fig. 4.24):
  \[ \langle E_{XZ}^{\text{BGO}} \rangle_{E_{XZ}^{\text{layer, max}}} > 0.2 \quad \langle E_{YZ}^{\text{BGO}} \rangle_{E_{YZ}^{\text{layer, max}}} > 1.8 \]  
  (4.16)

With these cuts, a total of 380563 events have been pre-selected and analyzed from the meta-classifier applying a threshold at 90% signal efficiency. The meta-classifier selected a total of 1971 signal events, corresponding to an overall fraction of $5.2 \times 10^{-3}$ that corresponds to the exact order of magnitude expected from the $e^-/p$ fraction in this region of the energy spectrum. As a good confirm of the purity of the selected sample, the distributions of three
4.2 Deep Learning classification approach with on-orbit data

Figure 4.23: Pre-selection on deposited energy spectrum.

Figure 4.24: Pre-selection on containment variables.
major features showed for the Random Forest classifier are showed for these
real data observations and can be easily compared to that in Figures 4.12,
4.13 and 4.14

Figure 4.25: Distribution for the F-value feature at layer 14 for selected electrons
(green) and rejected protons (blue)
4.2 Deep Learning classification approach with on-orbit data

Figure 4.26: Distribution for the energy in layers 13 and 14 for selected electrons (green) and rejected protons (blue)

Figure 4.27: Distribution for the RMS$^2$ feature at maximum of the shower for selected electrons (green) and rejected protons (blue)
Conclusions

The BGO calorimeter is the key detector of the DAMPE satellite, and with its almost 32 $X_0$ of depth, puts this experiment in a leading role in the present space-based research of high energy cosmic rays. It allows to investigate a portion of the energy spectrum that ranges from 5 GeV to 10 TeV for electrons and gamma-rays observations.

Analyzing the beam test campaign at CERN, in the first part of this work the expected resolution of 1% at 100 GeV has been verified along with a good linearity in energy reconstruction after correcting for energy losses in the calorimeter structure. The sampled structure of the calorimeter allows to image the showering of particles interacting with the crystal scintillators with a good resolution.

The particle identification with the BGO calorimeter represents the major analysis in this work, and combines the powerful features of the detector with some of the most powerful machine learning techniques used in data-science. The increasing use of machine learning techniques in high energy physics increases the amount of information obtained from detectors. In fact, for the DAMPE BGO, the sampled structure allowed to perform a powerful pattern recognition beside a classification based on selected features that characterize electromagnetic cascades.

The contamination of about $7.5 \times 10^{-5}$ in the selected signal sample is very close to the background rejection of $10^5$ expected from the whole DAMPE
detector. Further improvements can be made adding features from the other sub-detectors, and with an increase of the statistics and accuracy of the training MC events.

The first application on on-orbit observed data proves the effectiveness of the classifier, giving a fraction of electromagnetic over hadronic component of observed particles of $5.3 \cdot 10^{-3}$ that corresponds to the expected order of magnitudes.

The extension of this analysis to the whole energy spectrum observed by DAMPE will require a careful study of the effectiveness at different energies of the features used in the Random Forest, and the search for others more effective in different zones of the energy spectrum, while the pattern recognition algorithm would find improvements in more sophisticated network architectures adapted to the particular structure of the calorimeter.
Appendix A

Electromagnetic cascades parameterization

The interaction of a high-energy electron or photon with a thick absorber results in an electromagnetic cascade as pair production and bremsstrahlung generate more electrons and photons with lower energy. The longitudinal development is governed by the high-energy part of the cascade, and therefore scales as the radiation length (denoted as $X_0$ and defined as the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung or the $7/9$ of the mean free path for pair production by a high-energy photon) in the material. Electron energies eventually fall below the critical energy $E_c$ (defined as the energy at which the bremsstrahlung loss rate and the ionization loss rate are equal), and then dissipate their energy by ionization and excitation rather than by the generation of more shower particles.

It is convenient to introduce the following scale variables:

$$t = x/X_0, \quad y = E/E_c$$

(A.1)

Longitudinal profiles of electromagnetic cascades from an EGS4 simulation of a 30 GeV electron-induced cascade in iron are shown in Fig[A.1][2].
Figure A.1: An EGS4 simulation of a 30 GeV electron-induced cascade in iron. The histogram shows fractional energy deposition per radiation length, and the curve is a gamma-function fit to the distribution. Circles indicate the number of electrons with total energy greater than 1.5 MeV crossing planes at $X_0/2$ intervals (scale on right) and the squares the number of photons with $E \geq 1.5 \text{MeV}$ crossing the planes (scaled down to have the same area as the electron distribution.

The electron number in the distribution falls off more quickly than energy deposition. This is because, with increasing depth, a larger fraction of the cascade is carried by photons.

The mean longitudinal profile of the energy deposition in an electromagnetic cascade is well described by a gamma distribution:

$$\frac{1}{E_0} \frac{dE}{dt} = \frac{\beta t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}$$  \hspace{1cm} (A.2)
The maximum \( t_{\text{max}} \) occurs at

\[
    t_{\text{max}} = \left( \frac{\alpha - 1}{\beta} \right) = \ln(y) + C_e \quad j = e, \gamma
\]  

(A.3)

where \( C_e = -0.5 \) for electron-induced cascades and \( C_\gamma = +0.5 \) for photon-induced cascades. For many purposes it is sufficient to take \( \beta \approx 0.5 \).

The transverse development of electromagnetic showers in different materials scales with the Molière radius \( R_M \) defined as:

\[
    R_M = X_0 \frac{E_s}{E_c}
\]  

(A.4)

where \( E_s \sim 21 \text{ MeV} \). In a material containing a weight fraction \( w_j \) of the element with critical energy \( E_{cj} \) and radiation length \( X_j \), the Molière radius is given by

\[
    \frac{1}{R_M} = \frac{1}{E_s} \sum_j w_j E_{cj} \frac{1}{X_j}
\]  

(A.5)

On the average, only 10\% of the energy lies outside the cylinder with radius \( R_M \). About 99\% is contained inside of 3.5 \( R_M \), but at this radius and beyond composition effects become important and the scaling with \( R_M \) fails.


