The analysis strategy for the measurement of the electron flux with CALET on the International Space Station

E. Berti, a, b, *, L. Pacini b and Y. Akaike c on behalf of the CALET Collaboration
(a complete list of authors can be found at the end of the proceedings)

a University of Firenze, Via G. Sansone 1, Sesto Fiorentino, Firenze, Italy
b INFN Section of Florence, Via B. Rossi 2, Sesto Fiorentino, Firenze, Italy
c University of Waseda, 1 Chome-104 Totsukamachi, Shinjuku City, Tokyo, Japan
E-mail: eugenio.berti@fi.infn.it

The CALorimetric Electron Telescope (CALET), operating aboard the International Space Station since October 2015, is an experiment dedicated to high-energy astroparticle physics. The primary scientific goal of the experiment is the measurement of the electron+positron flux up to the multi-TeV region, which can provide unique information on the presence of nearby astrophysical sources and possible signals from dark matter. Other important goals are the ones relative to the flux of nuclear species from proton to iron up to tens of TeV/nucleon and to gamma-ray astronomy up to a few TeV. In order to accomplish these tasks, the CALET instrument was carefully designed exploiting a calorimeter solution composed by three detectors: CHarge Detector (CHD), IMaging Calorimeter (IMC) and Total AbSorption Calorimeter (TASC). This geometry allows for an excellent electromagnetic shower energy resolution (2%), a very high proton rejection factor ($10^5$) and a relatively large geometric factor (0.1 m$^2$sr). In this contribution, we present the analysis strategy employed for the measurement of the electron+positron flux, which is divided in two main steps. The first step consists of a group of selections to obtain a sample of well reconstructed candidates, removing particles outside the detector acceptance and particles with a charge $Z>1$, while keeping a high selection efficiency for electrons. The second step consists of a final rejection to remove the residual proton background: this is the most crucial point of the analysis since in cosmic rays protons are more abundant than electrons by a factor 100-1000. Proton rejection is performed using two different methodologies. We will demonstrate that, at low energies, it is enough to use a simple single cut that makes use of the reconstructed longitudinal and lateral profile, whereas, at high energies, it is necessary to use a more powerful cut that combines all detector information by the use of a multivariate analysis technique. Finally, we will show that this rejection algorithm leads to very stable performances at all energies, strongly reducing the impact of the associated uncertainty, which is the main source of systematic uncertainty in the high energy region.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter
1. Introduction

Since the first measurement carried out by the PAMELA experiment [1] and later confirmed by the AMS-02 [2] experiment, the unexpected excess in the cosmic ray positron flux raised a large interest due to its possible implication with the presence of nearby astrophysical sources and possible signals from dark matter. In addition, due to their strong energy loss during the diffusion in the Galaxy, it is expected that TeV electrons and positrons observed at Earth are likely originated in sources younger than $10^5$ years, and nearer than 1 kpc. In order to shed light on these processes, it is necessary to extend these measurements at higher energies, but, due to the limited mass, power and volume budget of a space mission, it is very difficult to accomplish this task with an instrument based on a magnetic spectrometer. On the other side, the electron+positron flux can be measured up to the multi-TeV region by exploiting an instrument based on a large, homogeneous and finely-segmented calorimeter. The CALorimetric Electron Telescope (CALET) [3] is the first space experiment dedicated to the measurement of cosmic rays to be based on this new design concept. The instrument, installed on the Exposed Facility of the Japanese Experiment Module (JEM-EF) of the International Space Station (ISS), started its operations in October 2015. After the originally scheduled 5 years of data taking, in 2021 the operations were extended for other 3 years, up to 2024. This decision was motivated by the relevant results obtained so far by the experiment and their expected improvement with an increased statistical sample. Up to 2021, the CALET collaboration published several results, relative to the measurement of the electron flux up to 4.8 TeV [4], the proton flux up to 10 TeV [5], carbon and oxygen up to 2.2 TeV/nucleon [6] and iron up to 2 TeV/n [7]. In addition to astroparticle physics, the CALET instrument allows to perform gamma-ray astronomy by the use of the calorimeter [8] and gamma-ray burst detection thanks to the Calet Gamma-ray Burst Monitor (CGBM) [9].

2. The experiment

The calorimeter constitutes the main detector of the CALET instrument and, as shown in Fig.1, it is composed by three detectors: CHarge Detector (CHD), IMaging Calorimeter (IMC), and Total AbSorption Calorimeter (TASC). The CHD is a charge detector composed of a couple of X and Y layers, each one made by 14 plastic scintillator paddles. Each paddle has size of $448 \, \text{mm} \times 32 \, \text{mm} \times 10 \, \text{mm}$, read-out using a single PhotoMultiplier Tube (PMT). The purpose of the CHD is to reconstruct the charge of the incident particles from $Z=1$ to $Z=40$ with excellent resolution. The IMC is a sampling and imaging calorimeter composed of 8 alternated couples of X and Y layers, each one made by 448 scintillating fibers (SciFi), interleaved with thin tungsten sheets for a total thickness of 3 $X_0$. The SciFis, each one having length of 448 mm and section of 1 mm, are read-out using 64-channel Multi-Anode PMTs (MAPMTs). The purpose of IMC is to reconstruct the direction of the incident particles with a resolution better than $0.5^\circ$ for high energy electrons and to sample the early stage of the shower development in the calorimeter. The TASC is a homogeneous and segmented calorimeter composed of 6 alternated couples of X and Y layers, each one made of 16 PWO scintillator crystal logs for a total thickness of 27 $X_0$. Each log has size of $20 \, \text{mm} \times 19 \, \text{mm} \times 320 \, \text{mm}$ and is read-out in the following way: by PMTs for the logs in the first layer, and by Avalanche PhotoDiode (APD) / Photodiode (PD) pairs for the logs in the remaining
The analysis strategy for the measurement of the electron flux with CALET

E. Berti

Figure 1: Event view corresponding to a 3.05 TeV electron candidate. The color scale represents the amount of energy deposit in each unit of the three detectors of the CALET calorimeter: CHD, IMC, and TASC.

eleven layers. The reason for this choice is that the first layer is also used to generate the trigger logic and a fast response is therefore needed, while the APD/PD system (each one coupled to two front-end circuits with different gains) allows for the required high dynamic range. The purpose of TASC is to reconstruct the energy of the incident particle with a resolution better than 2% for electrons and to sample the shower profile needed to perform electron/proton discrimination with a proton rejection factor of about $10^5$. The geometric factor of the instrument is approximately $0.1 \text{ m}^2\text{sr}$: about half of the acceptance corresponds to events entering from the top layer of CHD and exiting from the bottom layer of TASC; the remaining half requires that the particle enters at most before the fifth X-Y layer of the IMC and that the track crosses at least 26.42 cm in TASC.

3. The analysis strategy

The extension of the measurement of the electron+positron flux at higher energies faces two main challenges: the limited statistical sample (since the flux strongly decreases following a power law function) and the large proton contamination (since in cosmic rays protons are more abundant than electrons by a factor 100-1000). For this reason, the analysis strategy is defined in a different way according to the energy range. Below 500 GeV, the statistics are very large, the proton contamination is relatively small, and the error bands are dominated by the systematic uncertainty: for these reasons, we use only about half of the full acceptance to select very well reconstructed events and we perform proton rejection using a simple single cut based on the longitudinal and lateral development of the shower in TASC. On the opposite side, above 500 GeV, the statistics are very limited, the proton contamination is significantly large, and the error bands are dominated by the statistical uncertainty: for these reasons, we use the full acceptance to increase the size of the useful statistical sample and we perform proton rejection using a complex multivariate approach based on several variables from the shower development in IMC and TASC.

The analysis strategy is divided in two main steps. The first one, described in Sec. 3.1, involves a group of selections to obtain a well reconstructed sample of electron candidates, removing contamination from events outside acceptance and particles with charge $Z>1$. The second one,
The analysis strategy for the measurement of the electron flux with CALET described in Sec. 3.2, involves a proton rejection cut to further suppress the proton background, carried out with a different algorithm at low and high energies. All the variables used in the analysis were carefully validated to ensure a significant background rejection while preserving a very high electron selection efficiency and, at the same time, a good agreement of their distributions between Monte Carlo simulations and flight data. In particular, the analysis was carried out with two independent simulation data sets, generated using the EPICS \cite{10} and GEANT4 \cite{11} simulation frameworks, which have slightly different properties. For example, regarding the electromagnetic showers, EPICS reproduces better the transverse profile in IMC, whereas GEANT reproduces better the longitudinal profile in TASC. These differences were taken into account by running the analysis using slightly different variables in the two cases, in order to avoid possible model bias. In the following, we describe the variables used to measure the flux employing the EPICS simulation.

3.1 Selections

The group of selections aims to obtain a well reconstructed sample of electron candidates, removing contamination from events outside acceptance and particles with charge $Z>1$, while keeping a very high electron selection efficiency. In the following, these selections are listed in the same order they are applied in the analysis, together with a short description:

- **Offline trigger confirmation** - This selection requires a energy deposit larger than 50 MIP in the last two layers of IMC and larger than 100 MIP in the first layer of TASC. This selection corresponds to the same condition of the hardware trigger, but with an increased value of the thresholds, in order to select a flat region of the trigger discriminator efficiency.

- **Geometrical condition** - This selection requires a track reconstructed in one of the acceptance regions defined for the analysis (different below and above 500 GeV).

- **Track quality** - This selection requires a well reconstructed track, which is reconstructed using an algorithm developed for electromagnetic showers \cite{12}.

- **Charge selection** - This selection requires that the quadratic mean of the energy deposit in the paddles crossed by the reconstructed trajectory is smaller than 3.5 MIP. This selection removes most of background events from Helium and heavier nuclei, which are completely eliminated by the following cuts that suppress hadronic shower candidates.

- **Longitudinal shower likelihood** - This selection requires a longitudinal shower development in IMC and in TASC compatible with the electromagnetic case. A threshold condition is applied to a likelihood estimator built by the use of simulations.

- **Lateral shower concentration** - This selection requires a lateral shower development in IMC compatible with the electromagnetic case. This is obtained by selecting events that, in the last IMC layer, deposit more than 40% of the energy release in that layer within 9 SciFis from the reconstructed trajectory (1 Moliere radius in tungsten).

As shown in Fig. 2, the thresholds were carefully studied so that, above 30 GeV, selection efficiency is higher than 95% for electrons and smaller than 1% for protons.
The analysis strategy for the measurement of the electron flux with CALET

E. Berti

3.2 Proton rejection

Due to the large amount of protons with respect to electrons in cosmic rays, proton background is still very high after all the selections described in the previous section. To further suppress this contribution we make use of a dedicated proton rejection cut, carried out using a simple single cut below 500 GeV and a multivariate approach above 500 GeV. In both cases, after building the corresponding distributions, we set the threshold in order to fix the electron selection efficiency to 80%. The single cut is based on the variable $K = \log(F_E) + 0.5 \times R_E [\text{cm}]$ with combines $F_E$, the fraction of the energy deposited in the last TASC layer respect to the total, and $R_E$, the second moment of the lateral energy deposit distribution respect to the shower axis in the first TASC layer:

$$F_E = \frac{\sum_{j=1}^{16} \Delta E_{12,j}}{\sum_{i=1}^{12} \sum_{j=1}^{16} \Delta E_{i,j}}$$

$$R_E = \sqrt{\frac{\sum_{j=1}^{16} \Delta E_{1,j} (x_{1,j} - x_{\text{track}})^2}{\sum_{j=1}^{16} \Delta E_{1,j}}}$$

(1)

where $x_{i,j}$ and $\Delta E_{i,j}$ represent the central position and the energy deposit of log $j$ from layer $i$. This single cut is based on the more contained longitudinal and lateral development of electromagnetic showers with respect to hadronic showers. Despite the simplicity it is very powerful, leaving a residual proton contamination after rejection below 5%, later subtracted to get the final flux.

The multivariate analysis approach is based on Boosted Decision Tree (BDT) algorithm, which resulted to have the best performances for our application among the different multivariate methods that we tested. The parameters of the algorithm were carefully optimized in order to have excellent discrimination and stable performances: in particular, we opted for a BDT made of 100 trees, each one with a depth of 20. The BDT estimator is constructed using 9 variables: $F_E$, $R_E$, 3 variables from the fit of the longitudinal profile in IMC, and 4 variables from the fit of the longitudinal profile in TASC. The function used to fit the longitudinal profile in IMC and TASC are, respectively:
The analysis strategy for the measurement of the electron flux with CALET

The 3 variables used for the BDT estimator from IMC fit are: \( p_0, p_1 \), and the goodness of fit in IMC \( \chi^2_{IMC} \). The 4 variables used for the BDT estimator from the TASC fit are: the shower maximum position \( \alpha / b \), the shower attenuation constant \( b \), the 5% shower depth \( t_{5\%} \), and the goodness of fit in TASC \( \chi^2_{TASC} \). After rejection, the residual proton contamination is smaller than 5% below 1 TeV and than 20% above 1 TeV, and is later subtracted to get the final flux. In a further optimization of the algorithm that is currently under study, the BDT estimator is constructed using 4 additional variables: the sum of deposit in CHD, the lateral shower concentration in IMC, the maximum ratio between the release in two adjacent IMC layers among the first 6 layers, and the ratio between the deposit in the 7th and 8th IMC layers. In this case, the training is carried out using a single energy bin above 500 GeV, and compensating this fact by properly applying energy-dependent correction factors for each variable. Thanks to this optimized rejection algorithm, it will be possible to extend the measurement of the flux up to 8 TeV, while keeping a residual proton background below 10%.

\[
\begin{align*}
\frac{dE}{dt}_{IMC} &= e^{p_0+p_1 t} \\
\frac{dE}{dt}_{TASC} &= E_0 \frac{\beta + 1}{\Gamma(\alpha + 1)} t^{\alpha} e^{-bt}
\end{align*}
\]
The analysis strategy for the measurement of the electron flux with CALET

E. Berti

The analysis strategy for the measurement of the electron flux with CALET

E. Berti

The stability of BDT against the defined electron selection efficiency is estimated by varying it at 1% step between 70% and 90%, using 100 different training and test samples for each value of the efficiency. Putting together these 100 different samples for each of the 21 different values of the efficiency, we construct a histogram with the results of these different trials for each energy bin. Fig. 3 shows the BDT estimator distribution and the corresponding BDT stability for the energy bin $753.6 < E \text{ [GeV]} < 1194.3$. The careful optimization of the algorithm leads to an excellent agreement between the distributions in simulations and in data, and to a very stable result against different selection efficiencies and different training/test samples. The systematic uncertainty due to BDT stability, estimated by using the standard deviation (and the mean) of the histograms just described, is shown in Fig. 4. Due to the limited statistics, the BDT algorithm is less stable above 1 TeV, where this effect represents the dominant contribution to the systematic uncertainty.

4. Conclusions

The main goal of the CALET experiment is the measurement of the electron+positron flux up to the multi-TeV region. For this purpose, the collaboration developed an analysis procedure that is able to extend the measurement at such a high energy, both maximizing the effective acceptance and minimizing the proton contamination. Above 500 GeV, this is possible by using the full detector acceptance (geometric factor $\sim 0.1 \text{ m}^2\text{sr}$) and an optimized BDT-based proton rejection (residual background $< 10\%$). The parameters of the BDT algorithm were accurately studied in order to have stable performances with respect to a change in the electron selection efficiency and in the training/test samples. Due to the limited statistics, BDT stability represents the dominant contribution to the systematic uncertainty above 1 TeV. Thank to this well-established analysis procedure, in 2017 the CALET collaboration published the measurement of the electron+positron flux up to 4.8 TeV. With the increased statistics obtained in the following years of operations and the
optimization of the BDT algorithm described here, it will be possible to improve this measurement and extend it to higher energies.

References


The analysis strategy for the measurement of the electron flux with CALET