Fermi-LAT Measurement of Cosmic-ray Proton Spectrum
Paper Outline - Version 0

David M. Green

February 21, 2016

Abstract

The Pass 8 gamma-ray simulation and reconstruction package for the Large Area Telescope (LAT) on the Fermi Gamma-ray Space Telescope has allowed for the development of a new cosmic-ray proton analysis. Using the Pass 8 direction and energy reconstruction, we create a new proton event selection. This event selection has an acceptance of 1 m^2 sr over the incident proton energy range from 50 GeV to over 8 TeV and when applied to over 7 years of LAT observations provides over 700 million events for a spectral measurement. The systematic errors in the acceptance and energy reconstruction require careful study and will contribute significantly to the spectral measurement. The event selection and spectral measurement of the Pass 8 proton analysis opens the door to additional proton analyses with the LAT, such as the evaluation of proton anisotropy. We present a detailed study on the measurement of the cosmic-ray proton spectrum with Pass 8 data for the Fermi LAT.

1 Introduction

(I) Describe overview of LAT
- (A) Launch date to give early context of how much data there is available
- (B) Orbital parameters to show what kind of space environment we have to deal with
- (C) Development of Pass 8, short list of improvements and how this enables use to make a new proton analysis with the LAT

(II) Discuss recent developments of the CR proton spectrum from instruments
- (A) AMS-02 observes break in spectrum at 300 GeV
- (B) Potentially resolves discrepancy between satellite measurements in 100s GeV energy and balloon-borne measurements
- (C) But.... AMS-02 only goes to 1.8 TeV, statistics limited due to small acceptance and X years of flight
- (D) Gap left between 1.8 TeV of AMS-02 and 3 TeV of CREAM

(III) Goals of this analysis
- (A) Measure the cosmic-ray proton spectrum from 50ish GeV to several TeV
- (B) Fermi LAT in unique position to measure spectrum spanning between satellite measurements and balloon borne measurements
- (C) Also able to confirm spectral break as currently only seen by AMS-02 and possibly by Pamela
- (D) Create a new data set of cosmic-ray protons for future analysis (I'm not sure we really need this in the paper but might be nice to mention)

(IV) Event selection for high quality proton sample

(V) Energy reconstruction, biases, energy resolution, and limitations

(VI) Describe out instrument response: acceptance and contamination

(VII) Describe the methods used for spectral reconstruction: unfolding and forward folding using response matrix derived from MCs

(VIII) Describe evaluation of systematic uncertainties

- (A) Due to event selection: acceptance and contamination
- (B) Energy measurement: absolute energy scale and energy resolution
- (C) From hadronic model of Geant4 simulations
- (D) Spectral reconstruction: comparing unfolding and forward folding methods

(IX) Finally discuss observations and features of measured spectral, including possible spectral break and agreement with recent results (definitely need to but this in context with other measurements since while energy resolution is poor and systematics less precises than AMS-02 we can extend the energy further into the region of balloon-borne detectors which have never been done before and makes a quantitative connection between two different observation environments)

2 Event Analysis

(I) Overview

- (A) Description of the LAT
- (B) 4×4 array of towers which measure direction and energy of incoming cosmic-ray
- (C) Each tower is composed of TKR and CAL
- (D) TKR information
  - (i) Each TKR module is 18 x-y planes of silicon-strip detectors with tungsten converter foil
  - (ii) Total of 1.5 X_0 at normal incidence (should convert this to nuclear interaction length)
  - (iii) X-Y nature and depth of TKR allows for determination of initial direction of cosmic-ray
  - (iv) Additionally able to measure the time over threshold of CR
  - (v) ToT allows for measurement of signal $\propto \mathrm{Z}^2$
  - (vi) The last 4(?) Tungsten converter foils are thicker than the previous layers to ensure gamma-rays convert within TKR
- (E) CAL information
  - (i) CAL is homogeneous electromagnetic calorimeter
  - (ii) Each CAL module is 96 CsI(Tl) crystals in an hodoscopic array in 8 layers.
  - (iii) The hodoscopic nature of the CAL allows for measuring the shape and evolution of each particle shower which can be used with a profile fitter to determine the incident energy of the cosmic-ray
  - (iv) additional the imagine capability of the CAL allows for the measurement of the direction of the incident CR
  - (v) At normal incidence the CAL is 0.5 λ_i lengths deep but at horizontal incidence is it 1.5 λ_i deep
- (F) Anti-coincidence detector (ACD) surrounds the 4×4 tower array
- (G) ACD information
  - (i) 89 segmented covering 5 sides of the tower array
  - (ii) Each tile independently measures deposited energy from CR
  - (iii) Deposited energy $\propto \mathrm{Z}^2$

(H) Description of the LAT triggers and filters and point towards the paper with more information

(I) LAT was not designed for accurate measurement of hadronic showers

- (i) Very shallow homogeneous calorimeter not idea for fully capturing energy hadronic shower profile
- (ii) Compare to CREAM and/or AMS-02
- (iii) Unable to measure energy on an event by event basis, need to focus on a statistical ensemble approach with high event rate
- (iv) Therefore need to be aware of limitation of energy measurement and associated systematic uncertainties

(II) Pass 8 Event Reconstruction

- (A) I'm not 100% sure of the depth of this section but seeing as though we are using Pass 8 and that was a somewhat critical step into enabling this analysis's possibility I think having a dedicated section in the Event analysis chapter might make sense. If we put it anywhere it should be rather early before the simulations and after the describing the instrument
- (B) Pass 8 is the new event reconstruction and simulation software developed by the Fermi LAT collaboration that drastically improves LAT's performance
- (C) New event classification using boosted decision trees in TMVA
  - (i) Several new Pass 8 variables have been created to determine the quality of the direction, energy, and gamma-ray quality
- (D) More variables gives better separation between hadronic and leptonic showers in TMVA
- (E) Improved profile to fitting to particle showers improves energy measurement
  - (i) The New Full Profile fitter is able to extend the longitudinal profile of the shower outside the CAL therefore estimating the amount of energy leakage for high energy events, > 100 GeV
  - (ii) Two energies derived from new full profile fitter, one for TKR directions and one for CAL directions
- (F) New tree based TKR reconstruction allows for direction reconstruction at higher angles and larger energies
- (G) New ACD reconstruction provides better particle identification, lowering the contamination of proton sample
- (H) Is there something else I am missing from Pass 8? There is no Pass 8 paper to reference this so I am not sure how in depth I should go into this discussion.

(III) Monte Carlo simulations

- (A) Need to stress the importance of the simulations since this is how we derive all of our instrument response functions
- (B) Also use simulations for the development of TMVA selection to remove contamination for other CRs
- (C) Simulations based on Geant4
- (D) LAT instrument and spacecraft are fully simulated within Geant4
- (E) Particles with distributions of energies, directions, and charges are generated and propagated with realistic physics models for interactions with the simulated LAT which create raw data
- (F) Raw simulated data is processed through the same Pass 8 reconstruction software as flight data
- (G) We preform extensive comparison between simulated data and flight data to ensure results from MC analyses can be reliably applied flight data
- (H) Three types of simulations are used this analysis:
- (I) Proton simulation

(i) Simulation run from 4 GeV to 20 TeV
(ii) Cover 4π sr
(iii) Created with an $dN/dE \propto E^{-1.5}$ spectral index
(iv) Original purpose to study Pass 8 CR rejection for studying extragalactic background light
(v) This produces a simulation event sample of over X million events

(J) Electron simulation

(i) 10 GeV to 10 TeV
(ii) Cover 2π sr (the top half of the instrument)
(iii) Created with a $dN/dE \propto E^{-1.0}$ spectral index
(iv) Original purpose of studying instrument response for cosmic-ray electron analysis
(v) This produces a simulation event sample of over X million events

(K) Background simulation

(i) The background simulation was created to accurately simulate the cosmic-ray environment of the LAT during space flight
(ii) It contains CR particle from Z = 1 to Z = 26, electrons, positrons, neutrons, and Earth albedo gamma-rays
(iii) All particles are simulated with realistic fluxes using results from recent CR experiments
(iv) The background simulation used in this analysis simulates about 8 days worth of livetime
(v) Protons range: 4 GeV - 10 TeV and 4π sr
(vi) Electrons/Positrons range: 4 GeV - 10 TeV and 4π sr
(vii) Helium range: 4 GeV - 20 TeV and 4π sr
(viii) Heavier CR range: 2 GeV/amu - 50 GeV/amu and 4π sr
(ix) Fluxes are taken to be near solar minimum

(L) All simulations are produces with an additional setting called overlay events

(i) Overlay events are created from diagnostic events from flight data and signal is added on top of the simulated data
(ii) This is mimic the effect of having two events simultaneous enter the LAT (for such high events rates at lower energies is a reasonable assumption)
(iii) Pass 8 has many new algorithms to handle and reduce the effect of two simultaneous events interacting with the LAT

(IV) Event Selection

(A) Minimum Quality Cuts

(i) We want a selection of hight quality protons for the spectral analysis
(ii) TkrNumTracks > 0: Require an event to have at least one track
(iii) WP8CTPSFTail > 0.5: WP8CTPSFTail is a new Pass 8 TMVA variable which determines the quality of direction reconstruction, this ensures the track is well reconstructed
(iv) CalEnergyRaw > 20 GeV: This utilizes the high pass filter on the LAT. Any event with a deposited energy greater than 20 GeV is downloaded from the LAT. This means for events with CalEnergyRaw > 20 GeV is not effected by and lower energy filters which are difficult to understand for protons
(v) CalTrackAngle < 0.3: Ensure the difference between the CAL and TKR directions is small.
(vi) CalNewCfpSat: Ensure that the variables resulting new full profile fitter are not saturated
(vii) Tkr1LengthInCal > 200.0: Want a long path length through the calorimeter and does not fall within gaps of CAL. More active material will help ensure more deposited energy and a better reconstructed energy

- (viii) log10(TkrTree1ThickRLnNodes) < 1.0: To make sure events convert within the beginning of CAL and not last few layers of TKR (don't want to lose energy to TKR) we ensure the last few layers of the TKR do not have too many events This also helps with backslash and again losing energy back into the TKR
- (ix) These cuts ensure the direction and energy are well reconstructed, still need remove contamination source from helium + heavier CRs and electrons

(B) Helium and Heavier Ion Cut

- (i) To remove $|Z| > 1$ CRs we use two independent measures of charge in the LAT
- (ii) Tkr1TotTrAve
  - (a) The average time over threshold of signal in TKR
  - (b) Signal is ionization of CR interacting with silicon trips, therefore Signal $\propto Z^2$ in units of MIPs
- (iii) Acd2PLCTkr1TileActDistEnergy
  - (a) The path length corrected energy deposited in the ACD in units of MeV
  - (b) Signal once again due to ionization of CR interacting with plastic scintillator, therefore AcdPlcEnergy $\propto Z^2$
- (iv) Using these two measures of charge we create of phase-space of Tkr1TotTrAve vs Acd2PLCTkr1TileActDistEne
- (v) Designate a polygon in that phase-space around $\mathrm{Z}^2 = 1$ using BKG simulation
  - (a) It should be noted that Geant4 does not accurately recreate the rate of heavy ions interacting with the LAT, it greatly under estimates the event rate
  - (b) But since ionization is relatively simple and these heavy CRs aren't showering in the ACD or TKR, the amount of ionization energy is fairly accurate
  - (c) What this means is we can define the polygon and trust the position of the polygon in flight data
- (vi) We also include an additional cut on Acd2Cal1TriggerEnergy15
  - (a) Acd2Cal1TriggerEnergy15 is the energy deposited around a 15° cone from the CAL direction
  - (b) The reason for this is a population of large angle heavy CRs priests after the cut, they tend to have a poor TKR direction reconstruction but good CAL direction reconstruction. We demand the Acd2Cal1TriggerEnergy15 ¡ 30 MeV
- (vii) This greatly reduces the contamination of heavier ions and alphas to under 1% and decreasing at higher energies
- (viii) Also since ionization is $\propto Z^2$ we cannot differentiate protons from electrons using this cut

(C) Proton Classifier

- (i) To remove electrons from the proton sample we develop an event classifier with TMVA using protons as signal and electrons and background
- (ii) Since leptonic and hadronic showers are fundamentally different we can use several variables of merit from Pass 8 reconstruction to build a multi-variate analysis and differential protons from electrons
  - (a) The CALs ability to image showers means we have several variables with trace shower size and evolution (hadronic showers are much wider and longer than leptonic showers)
  - (b) Additional variables like centroid position of the shower, and size of the track left in the TKR also are able to distinguish between leptonic and hadronic showers
- (iii) Using simulations previously described with a boosted decision tree with TMVA to train classifier
- (iv) Using classifier output we can select on an optimal value for each energy bin and create an energy dependent event selection
  - (a) This is critical because the proton classifier's ability to distinguish protons from electrons is energy dependent

        (b) Classifier has trouble separating events at high energies (all events look very similar at highest energies since instrument is small compared to size of TeV showers)

    (v) The optimal value for each energy bin is selected by using the ROC curve (signal efficiency to background rejection) ensuring the derivative of the ROC curve is below a defined tolerance
    (vi) We also use several different proton scanning efficiencies
        (a) Find energy dependent cut on proton classifier output which ensures a predefined (90%,80%, 70%, etc...) is achieved
        (b) This allows us to probe systematic errors associated with the proton classifier, acceptance, and contamination
        (c) We produce a spectrum for each different scanning efficiency and compare it to the optimized selection
    (vii) We use a template fitter developed by the Fermi-LAT electron analysis to evaluate the data/MC agreement for the variables used in the training the classifier and the output of the classifier
        (a) Data/MC agreement is very important
        (b) MC needs to properly recreate data variables in order trust classifier is giving the desired results
        (c) Discrepancies between Data/MC in training variables will translate to systematic errors associated with classifier
        (d) Also dependent on how important the variable is for training the classifier
        (e) Data/MC agreement for classifier output is very good
        (f) Hopefully systematic uncertainties are low

(V) Energy reconstruction

    (A)

3 Spectral Analysis

(I) Instrument Acceptance

(II) Residual Contamination

(III) Spectral Reconstruction

(IV) Systematic Uncertainties

4 Results and Discussion