The Fermi Large Area Telescope

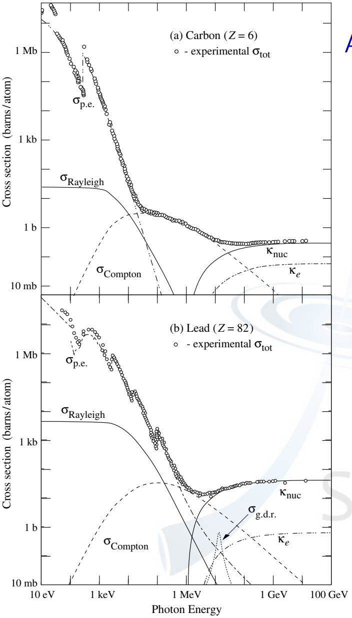
An overview

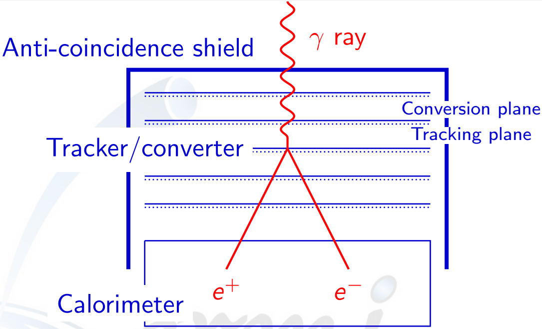
Luca Baldini
INFN–Pisa and University of Pisa
luca.baldini@pi.infn.it

Fermi Summer School 2012
Lewes, May 31, 2012

Outline

- Detection principle: the concept of a pair conversion telescope.
  - basic design drivers.
- The Large Area Telescope:
  - silicon tracker (TKR);
  - electromagnetic calorimeter (CAL);
  - anti-coincidence detector (ACD).
- Detection principle revisited.
- Orbital environment.
  - (And instrumental pile-up, aka *ghost* effect.)
- Event triggering and filtering.
- Event-level analysis:
  - event reconstruction;
  - background rejection.
- Conclusions.
- All is very IRF-oriented!
- There are a few Exercises for you to solve in the following slides.





- Pair production is the dominant interaction process for photons in the LAT energy range;
- e^+e^- pair provides the information about the γ-ray direction/energy;
- e^+e^- pair provides a clear signature for background rejection (really?).

Science design drivers

- Effective area and Point Spread Function:
  - thickness and layout of conversion layers;
  - PSF also drives the design of the sensors, the spacing of the detection planes and the overall TKR design.
- Energy range and resolution:
  - thickness and design of the calorimeter;
- Field of view:
  - determined by the aspect ratio of the instrument;
- Charged particle background rejection:
  - mainly drives the ACD design;
  - also impacts the TKR and CAL design (which are needed for the background rejection).
  - need for a flexible triggering and event filtering system.

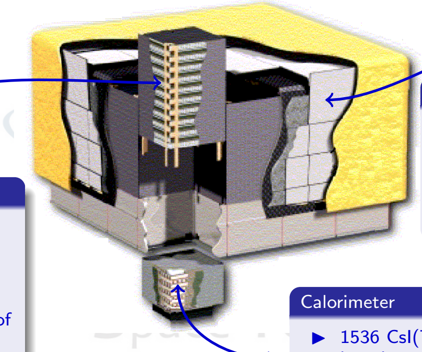
Mission design drivers

- Launcher type and allocated space:
  - maximum possible lateral dimensions of the instruments (i.e. geometric area);
  - about $\sim 1.8 \times 1.8$ m^2 for Fermi (the LAT footprint is actually $\sim 1.5 \times 1.5$ m^2).
- Power budget:
  - number of electronics readout channels in the tracker (i.e strip pitch, number of layers);
  - about 650 W overall for Fermi;
- Mass budget:
  - essentially limits the total depth of the calorimeter (once the footprint is fixed);
  - 3000 kg for Fermi.
- Telemetry bandwidth:
  - need onboard filtering.
- Launch and operation in space:
  - sustain the vibrational loads during the launch;
  - operate in vacuum, sustain thermal gradients.

The Large Area Telescope

Large Area telescope

- Overall modular design.
- 4×4 array of identical towers (each one including a tracker and a calorimeter module).
- Tracker surrounded by and Anti-Coincidence Detector (ACD).
- *"It uses less power than a toaster and we talk to it over a telephone line."* (Bill Atwood)



Tracker

- Silicon strip detectors, W conversion foils; 1.5 radiation lengths on-axis.
- $\sim$ 10k sensors, 73 m^2 of silicon active area, $\sim$ 1M readout channels.
- High-precision tracking, short dead time.

Anti-Coincidence Detector

- Segmented (89 tiles) as to minimize self-veto at high energy.
- 0.9997 average detection efficiency.

Calorimeter

- 1536 CsI(TI) crystal; 8.6 radiation lengths on-axis.
- Hodoscopic, 3D shower profile reconstruction for leakage correction.

Silicon Tracker/Converter (1/2)

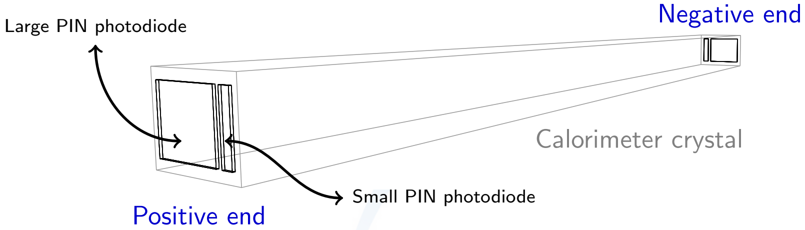
- Primary roles:
  - convert γ rays into electron/positron pairs;
  - main event trigger (more on this later);
  - direction reconstruction.
- Also important for:
  - background rejection (SSD veto, hit counting);
  - energy measurement at low energy (i.e., below a few hundred MeV).
- Use of Silicon Strip Detector (SSD) technology:
  - precise tracking with $\sim$ no detector-induced deadtime;
  - self-triggering.
- Key features:
  - ~ 73 m^2 of single-sided SSDs (400 μm thickness, 228 μm pitch);
  - 884,736 independent readout channels (~ 200 μW per channel);
  - digital readout (plus layer OR time over threshold);
  - $\sim 10^{-6}$ noise occupancy at the nominal 1/4 of a Minimum Ionizing Particle (MIP) threshold (providing $\sim 100\%$ detection efficiency).
- Exercise: Estimate the average number of noise hits per event in the full LAT.

- Tradeoffs in the design of the tracker converter:
  - overall thickness of the converter foils: conversion efficiency vs. multiple scattering (limiting the angular resolution at low energy);
  - number and spacing of the planes: energy dependence of the PSF;
  - strip pitch: hit resolution vs. power consumption.
- 18 paired x–y layers ($\sim$ 36 cm on a side, spaced by $\sim$ 3.5 cm) in two distinct sections:
  - front has better PSF and lower background contamination;
  - 1.5 X_0 on axis—that's a lot for a tracker!
- $\boxed{\text{Exercise:}}$ What's the maximum off-axis angle the TKR will trigger?

Electromagnetic Calorimeter (1/2)

- Primary roles:
  - energy reconstruction;
  - contribution to the event trigger (more on this later);
- Also important for:
  - background rejection (shower shape);
  - seeding the tracker reconstruction.
- Crystal detector elements:
  - 8 layers of 12 CsI(Tl) crystals ($27 \times 20 \times 326$ mm^3) per tower;
  - hodoscopic stacking (alternating orthogonal layers);
  - 8.6 X_0 on-axis.
- Readout electronics:
  - dual PIN photodiode on each crystal end;
  - each one processes by two electronics chains ($\times 1$, $\times 8$);
  - four readout ranges, dynamic range 2 MeV–70 GeV per crystal.
- Exercise: How much energy does a MIP on-axis release in the CAL?

Electromagnetic Calorimeter (2/2)

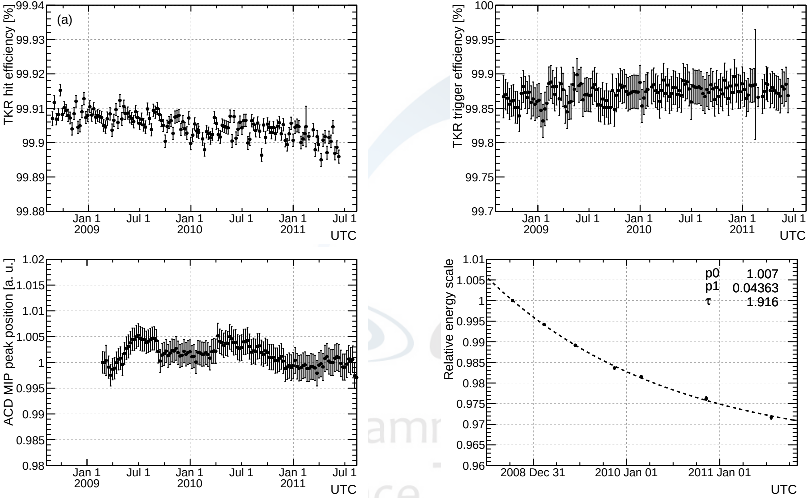


- CAL xtals with readout at each end:
  - measure longitudinal position of the energy deposition from light asymmetry;
  - provide a full 3-dimensional image of the EM shower;
- CAL imaging capabilities are crucial for both background rejection and energy reconstruction at high energy:
  - remember, the LAT is $\sim 10\ X_0$ on axis, so there is a significant shower leakage out the back of the CAL.
- $\boxed{\text{Exercise:}}$ What is the fraction of energy escaping out the back of the CAL for a 500 GeV photon on-axis?

Anticoincidence detector

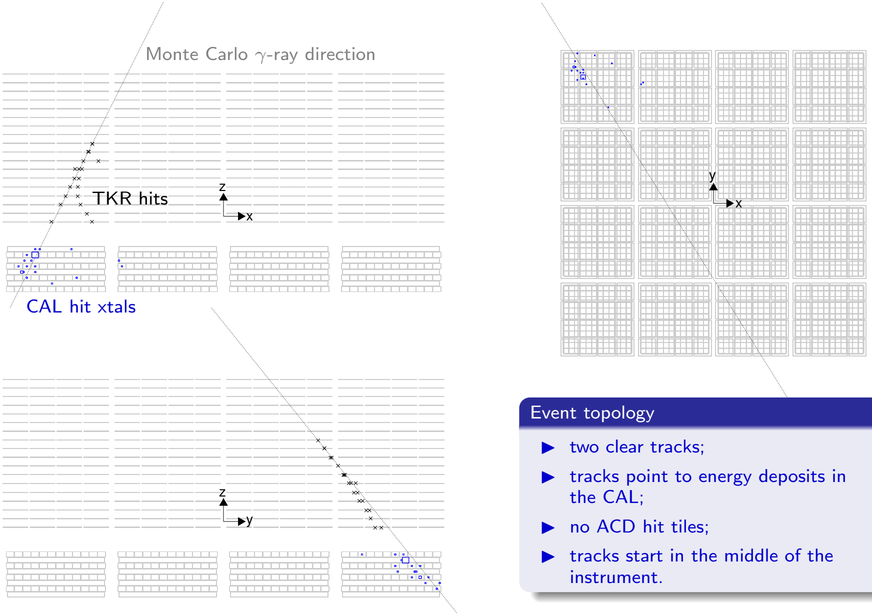
- Primary roles:
  - event triggering and onboard filter (more on this later);
  - background rejection.
- Also important for:
  - identifying heavy ions for CAL calibration purposes.
- One important lesson learned from the previous mission:
  - backsplash from the CAL in high-energy event can hit the ACD;
  - can cause *self-veto*, especially for monolithic shields.
- The LAT ACD is segmented:
  - 89 tiles (overlapping in one dimension) plus 8 *ribbons* (covering the gaps in the other);
  - can extrapolate tracks to specific tiles;
  - this also makes complete hermeticity more difficult to achieve.

Long-term trending and stability



- The LAT shows no significant degradation in time over the first $\sim$ three years of mission.
  - drift of the light yield in the CAL expected from radiation damage.

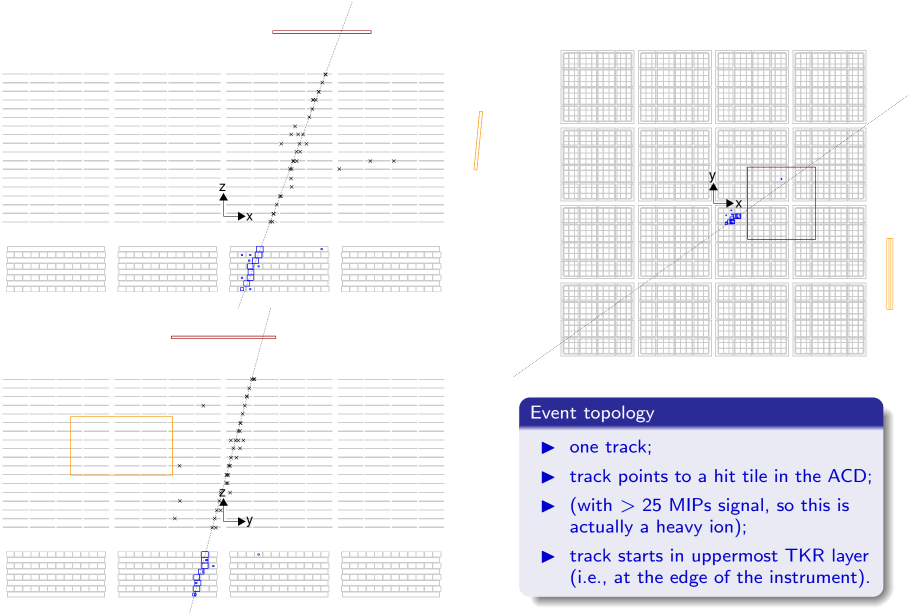
A "gold-plated" simulated 360 MeV γ-ray...



Event topology

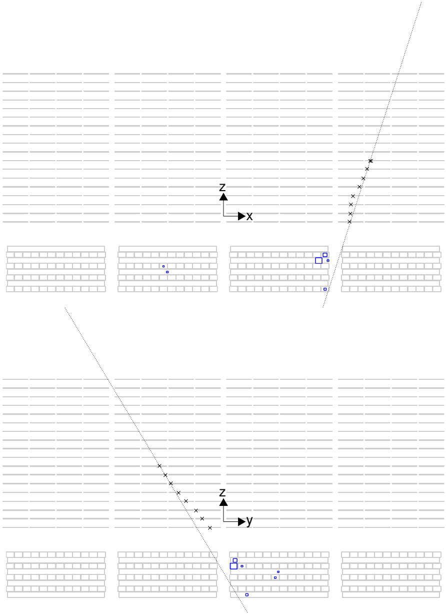
- two clear tracks;
- tracks point to energy deposits in the CAL;
- no ACD hit tiles;
- tracks start in the middle of the instrument.

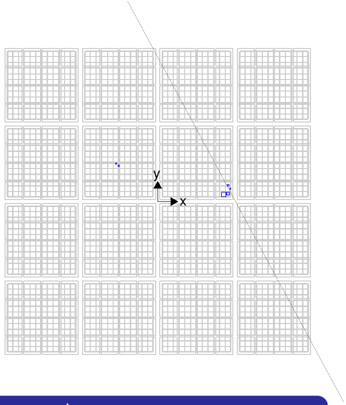
... AND A "GOLD-PLATED" BACKGROUND EVENT



Event topology

- one track;
- track points to a hit tile in the ACD;
- (with > 25 MIPs signal, so this is actually a heavy ion);
- track starts in uppermost TKR layer (i.e., at the edge of the instrument).

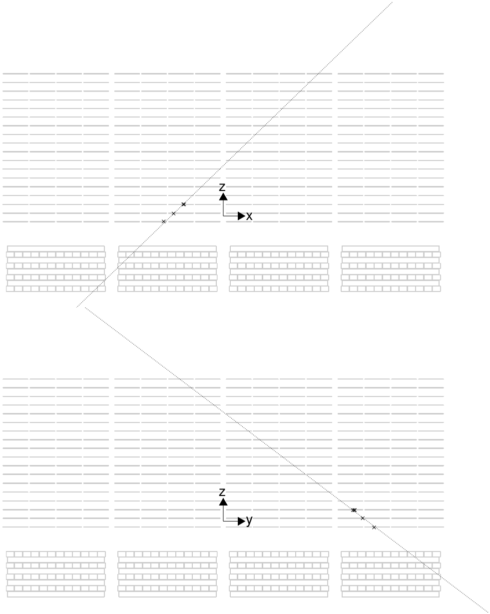


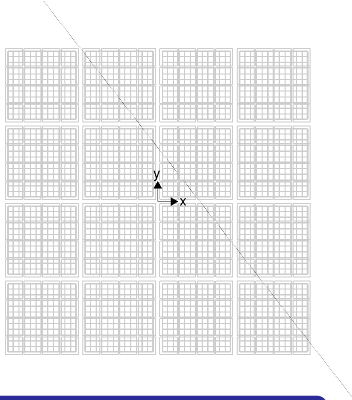


Event topology

- no ACD hit tiles (good);
- but this time we only have one track;
- where's the other guy (aren't we supposed to detect electron-positron pairs)?

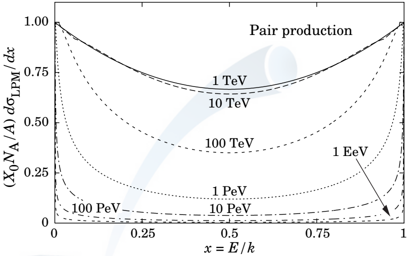
. . . AND A $\sim$ 40 MEV SIMULATED γ-RAY





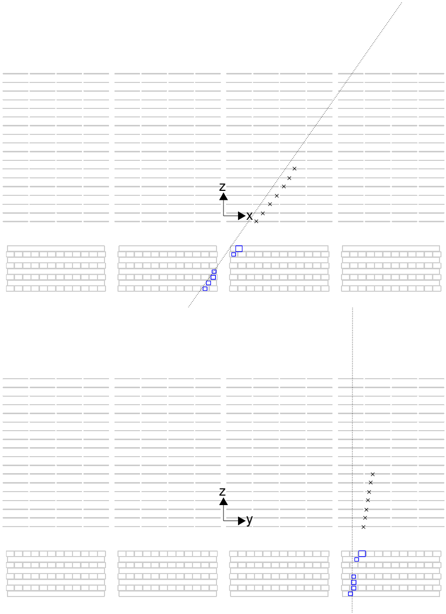
Event topology

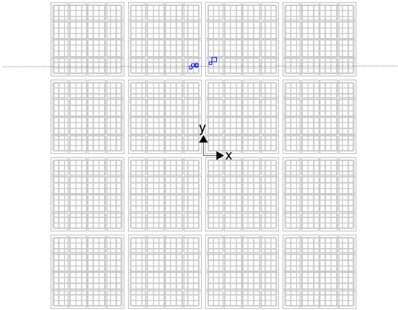
- no ACD hit tiles (good);
- we still only have one track;
- and even worse: it doesn't even make it to the CAL!
- Can we estimate the energy for this one?



- e^+ and e_- split the energy equally *on average*;
  - not uncommon that one takes the vast majority...
  - ...at the level that the other track can die in the tungsten.
- At high energy the opening angle is small:
  - at some point we don't resolve the two tracks anymore (back into the one-track case).
- Except for the stray tracks from CAL backsplash!
- Exercise: Give a rough estimate of the maximum energy at which the TKR is able to resolve the two tracks in the pair.

Back to the background

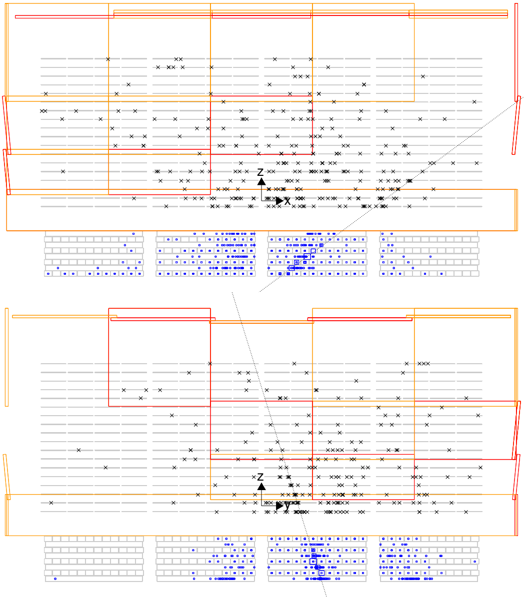


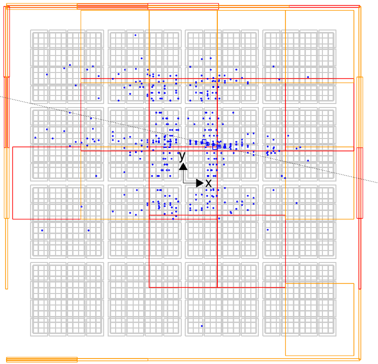


Event topology

- no ACD hit tiles (good);
- one track (can happen);
- corresponding energy deposit in the CAL (good);
- believe it or not this is a back-entering CR proton ☹

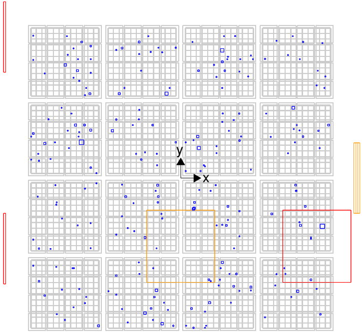
A simulated 540 GeV γ-ray





Event topology

- (potentially) many many tracks;
- lots of CAL back-splash;
- tracks point just about everywhere;
- many hits in the ACD;
- still we get energy and direction right (for this particular event).



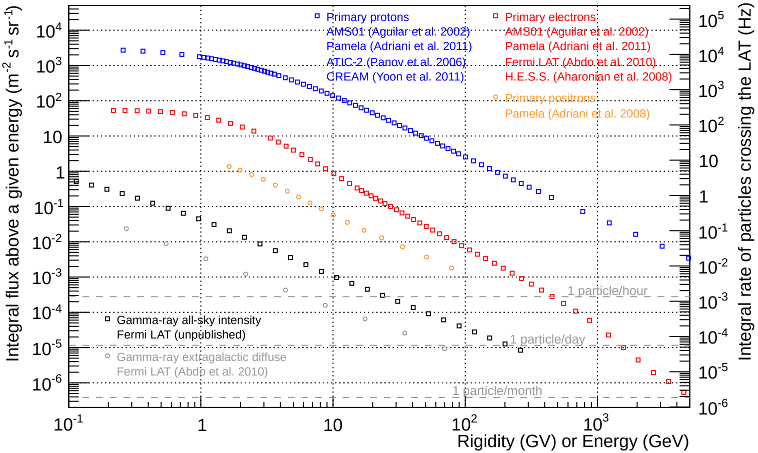
Event topology

- Pretty much the entire detector is on (many many low-energy photons);
- standard event reconstruction can't do much, here.
- (The instrument was in a special configuration/orientation; this would not pass the gamma onboard filter.)

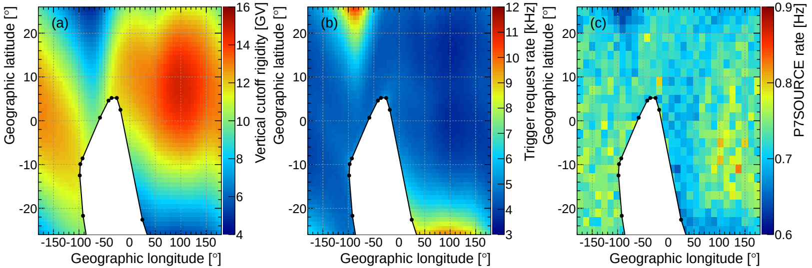
Now let's stop with event display but. . .

- γ-ray event topology varies *a lot* across the instrument phase space:
  - can have (zero), one, two or many tracks;
  - can have hits in the ACD (here is where we take advantage of the segmentation);
  - from "no energy deposit" to a "fully developed em shower" in the CAL.
- Background event topology varies *a lot* across the instrument phase space:
  - some of them are easy to identify;
  - some are hard;
  - some are impossible (e.g., the *irreducible* background).
- Take-away message 1: event reconstruction is challenging.
- Take-away message 2: background rejection is challenging.

Orbital environment (1/2)

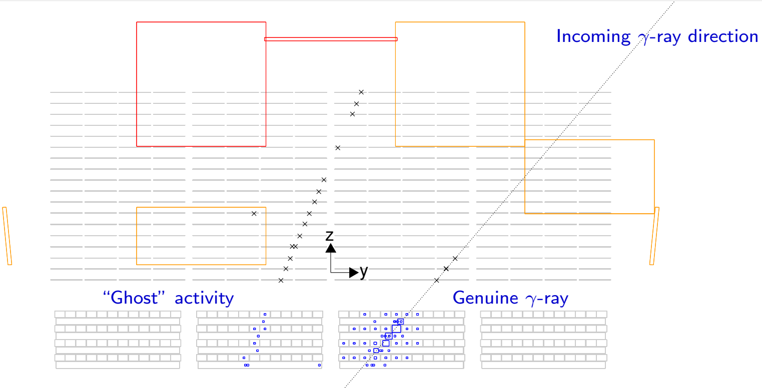


- Relevant γ-ray and primary CR spectra (not taking into account the effect of the geomagnetic field):
  - up to 10^6 background rejection power required;
  - running out of photons above a few TeV with ~ 2 m^2 sr acceptance.



- CR-induced background level in low-Earth orbit depends on the local geomagnetic conditions:
  - low-energy CRs effectively shielded by the Earth's magnetic field;
  - how low is low depends on the position (a).
- Most of the charged particles crossing the LAT generate a treq;
  - trigger request rate also varies across the orbit (b).
- Celestial γ-rays are unaffected by magnetic fields:
  - the rates of the cleanest event classes (c) should not depend on the local geomagnetic conditions; however. . .
- Exercise: Estimate the vertical rigidity cutoff at the equator.

Instrumental pile-up (aka "ghost" effect)



- The *persistence* time of the electronics signals in the detector is of the order of $\sim 10\ \mu$s:
  - if two events happen to be that close in time (and we happen to trigger on one) we're effectively reading out both.
  - Ghost signals can cause good γ rays to be misclassified as background (i.e., loss of effective area).
- $\boxed{\text{Exercise:}}$ estimate the fraction of events affected by ghosts.

Trigger and onboard filter basics

- Ideally we would like to be able to:
  - read out all the events (i.e., all particles crossing the detector);
  - down-link all the events to the ground;
  - postpone all the decisions (is the event a γ ray?) to the offline data analysis phase.
- Unfortunately that's generally impossible in high-energy physics experiments:
  - reading out an event takes time (*at least* 26.5 μs for the LAT); during this *deadtime* the instrument is *blind*;
  - the bandwidth for transmitting data to ground is limited ($\sim$ 1 Mb/s)—and expensive.
- Bottom line: we do have to take decisions onboard about:
  - which events we want to read out;
  - which events (among those that we read out) we want to transmit to ground.
- Exercise: estimate the deadtime fraction if we were to read out all the events causing a trigger request (take $\sim$ 8 kHz treq rate).

Triggering the LAT

- Use fast ($< 1\ \mu$s) signals to trigger readout;
  - as opposed to ground analysis using slower ($\sim 10\ \mu$s) signals.
- Each subsystem generates one or more *trigger primitives*:
  - `TKR`: three adjacent tracker x-y layers above threshold;
  - `CAL_LO`: any single CAL channel above 100 MeV (adjustable);
  - `CAL_HI`: any single CAL channel above 1 GeV (adjustable);
  - `ROI`: one or more ACD tile(s) over veto threshold (nominally 0.45 MIP) in proximity of a triggering TKR tower;
  - `CNO`: signal in any of the ACD tiles above the CNO (Carbon Nitrogen Oxygen) threshold (nominally 25 MIPs);
  - `PERIODIC`: 2 Hz synchronous (for minimum bias event sample).
- Some of the trigger primitives can open a 700 ns trigger window;
  - collect all the asserted primitives when the window is closed;
  - map each combination into a look-up table;
  - decide whether to read out the event or not.
- If the trigger request is accepted, the full LAT is read out.
  - (It takes $< 2\ \mu$s to take the decision.)
- Exercise: Does a MIP 45° off-axis generate a `CAL_LO`?

Trigger engines and prescales

Engine	PERIODIC	CAL_HI	CAL_LO	TKR	ROI	CNO	Prescale	Average rate [Hz]
3	1	×	×	×	×	×	0	2
4	0	×	1	1	1	1	0	200
5	0	×	×	×	×	1	250	5
6	0	1	×	×	×	0	0	100
7	0	0	×	1	0	0	0	1500
8	0	0	1	0	0	0	0	400
9	0	0	1	1	1	0	0	700
10	0	0	0	1	1	0	50	100

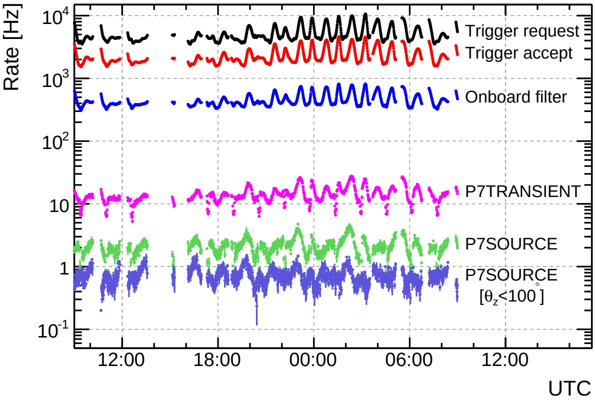
(1: required, 0: excluded, ×: either)

- Some trigger primitive combinations are prescaled.
- Consider trigger engine 10 for example:
  - `(TKR && ROI) && !(CNO || CAL_LO || CAL_HI)`
  - This is most likely to be a MIP and very unlikely to be a γ ray;
  - there are many of them: we only read out 1 every 50.
- Prescaling reduces deadtime:
  - we don't actually read out the event (which takes at least 26.5 μs).
- 5–10 kHz trigger request rate $\rightarrow$ 2–3 kHz event readout rate.
- Exercise: estimate the deadtime fraction for a 2.2 kHz readout rate.

Onboard filter

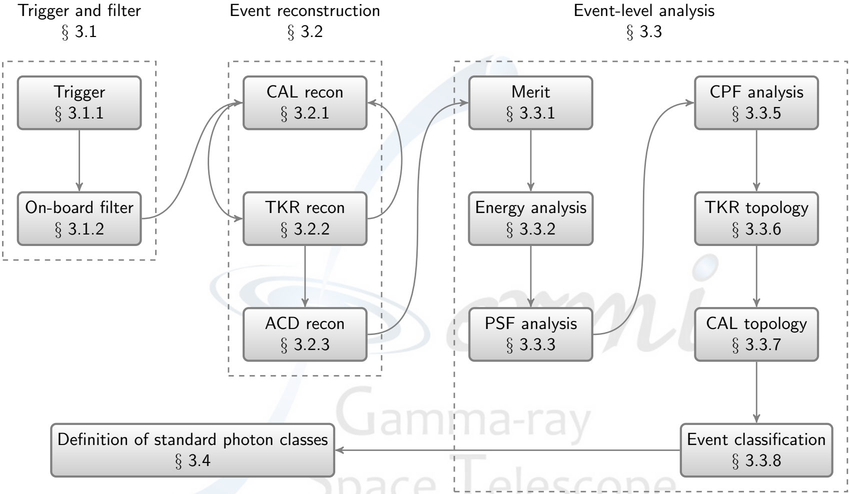
- We're down to 2.2 kHz average event readout rate;
  - with an average compressed event size of $\sim$ 500 bytes that's still too much;
  - need further onboard event filtering to reduce the rate of events to be transmitted to ground.
- Onboard filter: configurable, has access to the full event information;
  - hierarchical set of conditions with the fastest being applied first.
- Multiple coexisting filtering algorithms running:
  - gamma filter: keep whatever might possibly be a γ ray;
  - HIP filter: select heavy ion events for CAL calibration;
  - diagnostics filter: provide a prescaled unbiased sample of all trigger types.
- 2–3 kHz event readout rate $\rightarrow$ 300–500 Hz downlink rate.
- Exercise: estimate the necessary average downlink bandwidth for 2.2 kHz and 400 Hz event readout rate.

Data reduction overview



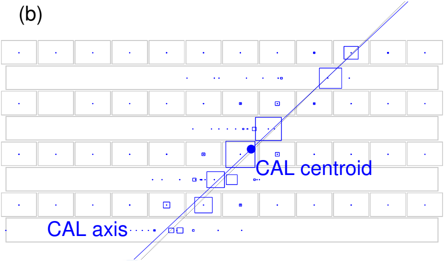
- Almost all the particles ($\sim$ 99%) downlinked to ground are still charged background.
  - (Though there is still interesting science in there.)
- The onboard filter is highly efficient for γ rays.
- The remaining data reduction steps are performed as part of the offline ground processing.

Event selection analysis overview



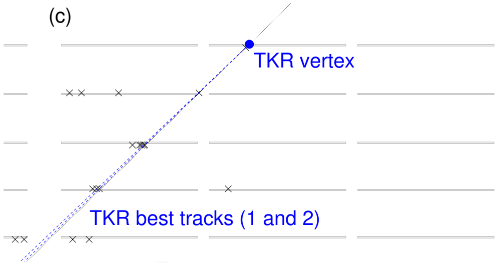
- (Disregard the section numbers for the purpose of this presentation.)

CAL reconstruction overview



- Apply xtal calibrations (i.e., convert ADC counts to MeV).
- Iterative moments analysis (i.e., calculate the principal axes of the inertia tensor associated with the energy deposition):
  - shower centroid;
  - shower direction ($\sim 1^\circ$ resolution above ~ 10 GeV);
  - shower transverse/longitudinal spread (background rejection).
- Energy reconstruction:
  - much, much more than summing up the xtal energies;
  - three different reconstruction algorithms;
  - different performance in different parts of the phase space.
- Note that we don't currently attempt to identify separate clusters of hit logs.

TKR reconstruction overview



- Combine adjacent hit strips to form clusters.
- Seed the track-finding stage with the CAL information, if available.
- Combinatoric search for tracks through a Kalman fit/filter technique:
  - start from a seed;
  - propagate to next plane based on the expected multiple scattering (need particle/energy hypothesis) and add hits as possible;
- Order tracks by quality (longest, straightest: best).
- Vertexing: combine the two best tracks when possible.
- (Much more complicated than this in real life.)

ACD reconstruction overview

- Apply tile/ribbon calibrations (i.e., convert ADC counts to MeV).
- Look for reasons to *veto* the event:
  - (i.e., decide it's a charged particle, as opposed to a γ ray).
- Much, much more complicated than requiring that there is no energy in the ACD:
  - a lot of phase space for weird things to happen;
  - (as you have seen before in the event displays).
- Extrapolate TKR tracks to the ACD:
  - is there any signal in the tile the track points to?
- But there are many ways we can potentially go wrong:
  - did we pick the right track?
  - did we happen to pass through inactive (or not fully efficient) areas in the ACD (i.e., ribbons, corners)?
  - are we affected by the backsplash (the energy deposited in the CAL is a good proxy for that).

- Complex multivariate analysis:
  - uses Classification Trees in conjunction with plain cuts;
  - a huge amount of work went into defining relevant classification variables.
- PSF analysis:
  - determine the best direction estimate (1^{st} track, vtx, neutral vtx);
  - along with a reconstruction quality indicator.
- Energy analysis:
  - select the best energy method (+ quality indicator).
- "Charged Particle in the Field of view" analysis:
  - identify events which are clearly charged particles in the FOV.
- TKR and CAL topology analysis:
  - probability of an event being a γ ray using CAL/TKR information.
- Event classification:
  - combine all the previous information.
- Definition of the photon classes.

Standard photon classes

- How different photon classes differ?
  - Primarily in the level of background contamination;
  - and, since you don't get anything for free, in the γ-ray efficiency too.
  - So we have (form dirtiest to cleanest) `P7TRANSIENT`, `P7SOURCE`, `P7CLEAN`, `P7ULTRACLEAN`.
- Why different photon classes?
  - Because different analyses require different signal-to-noise ratios.
  - (Or, phrased in a different way: different analyses might provide additional handles to reject background).
- Point source analysis:
  - cut on the ROI;
  - is it a pulsar? Even better, can use the pulse phase, too!
- Isotropic background:
  - no obvious spatial or temporal signatures to distinguish signal and background.
- $\boxed{\text{Exercise:}}$ how much background do you remove by selecting events in a 5° ROI?

Future developments in the event-level analysis

- The LAT provides a huge amount of information on an event-by-event basis.
- The instrument performance is not "once and forever":
  - you can improve by being more clever in the event reconstruction and in the background rejection;
  - even now that the LAT is built and up in space we can improve.
- Ongoing long-term effort to revisit all the aspects of the event-level analysis:
  - make use of the lessons learned operating the LAT;
  - new pattern recognition in the TKR;
  - clustering stage in the CAL;
  - new energy reconstruction at high energy;
  - new ACD reconstruction;
  - new event classification...

Conclusions

- The LAT is essentially a particle-physics instrument.
- Huge amount of information available on an event-by-event basis:
  - event reconstruction and background rejection can be very hard;
  - understanding the instrumental effects can also be very hard;
  - we always have room for improving the performance!
- Large dynamic range and field of view:
  - large variations in the event topology;
  - parametrizing the instrument response is challenging.
- There's a lot of stuff going on to get the photon energy, direction and arrival time from the raw detector information!