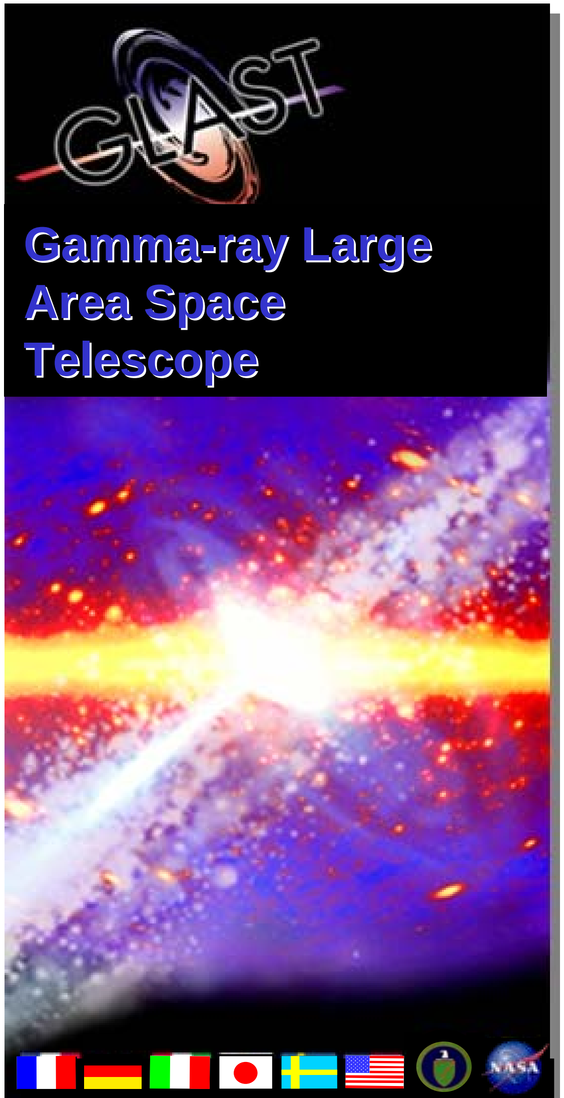


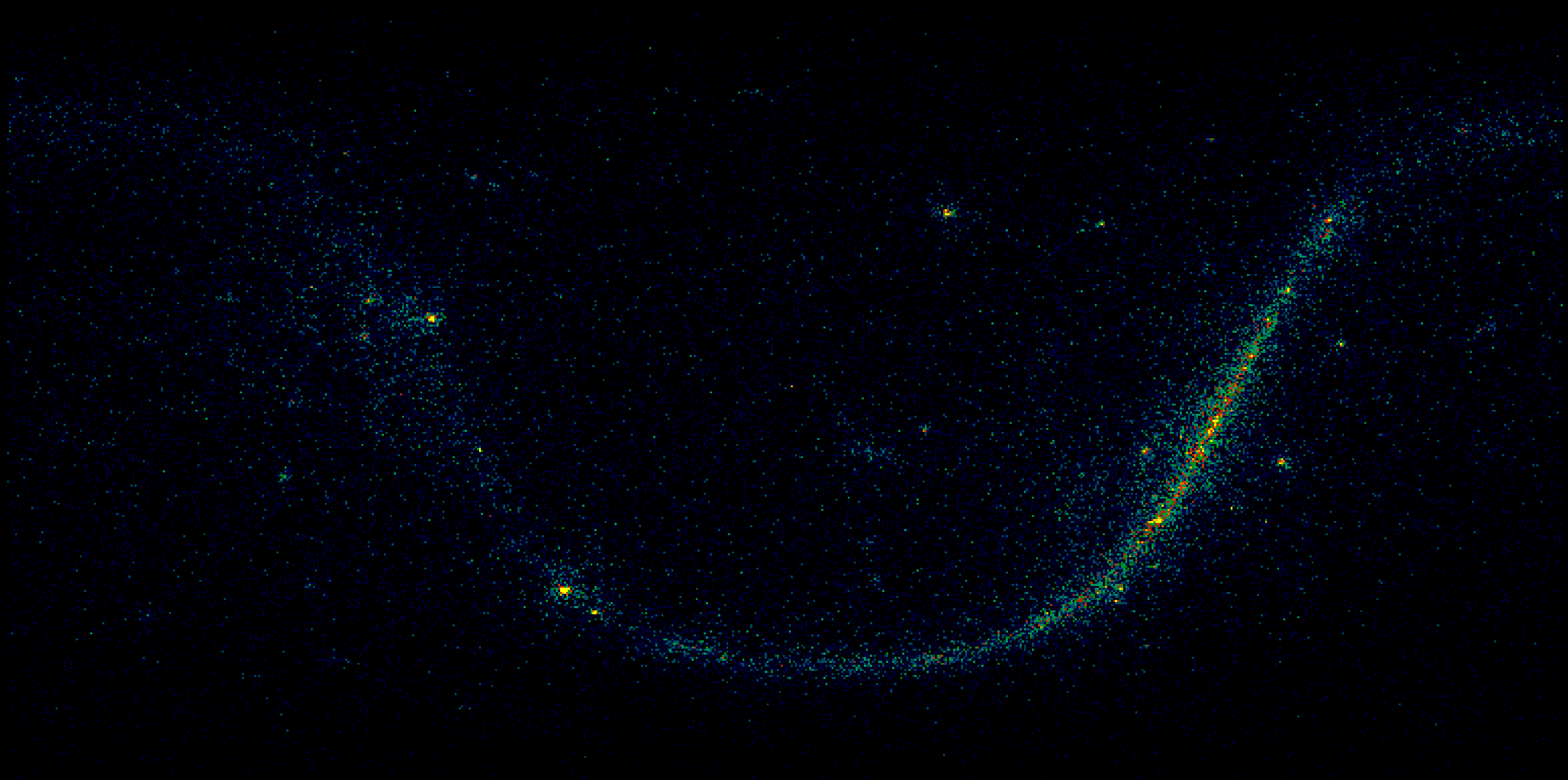
GLAST Large Area Telescope:

Data Challenge 2

Getting Ready for Science

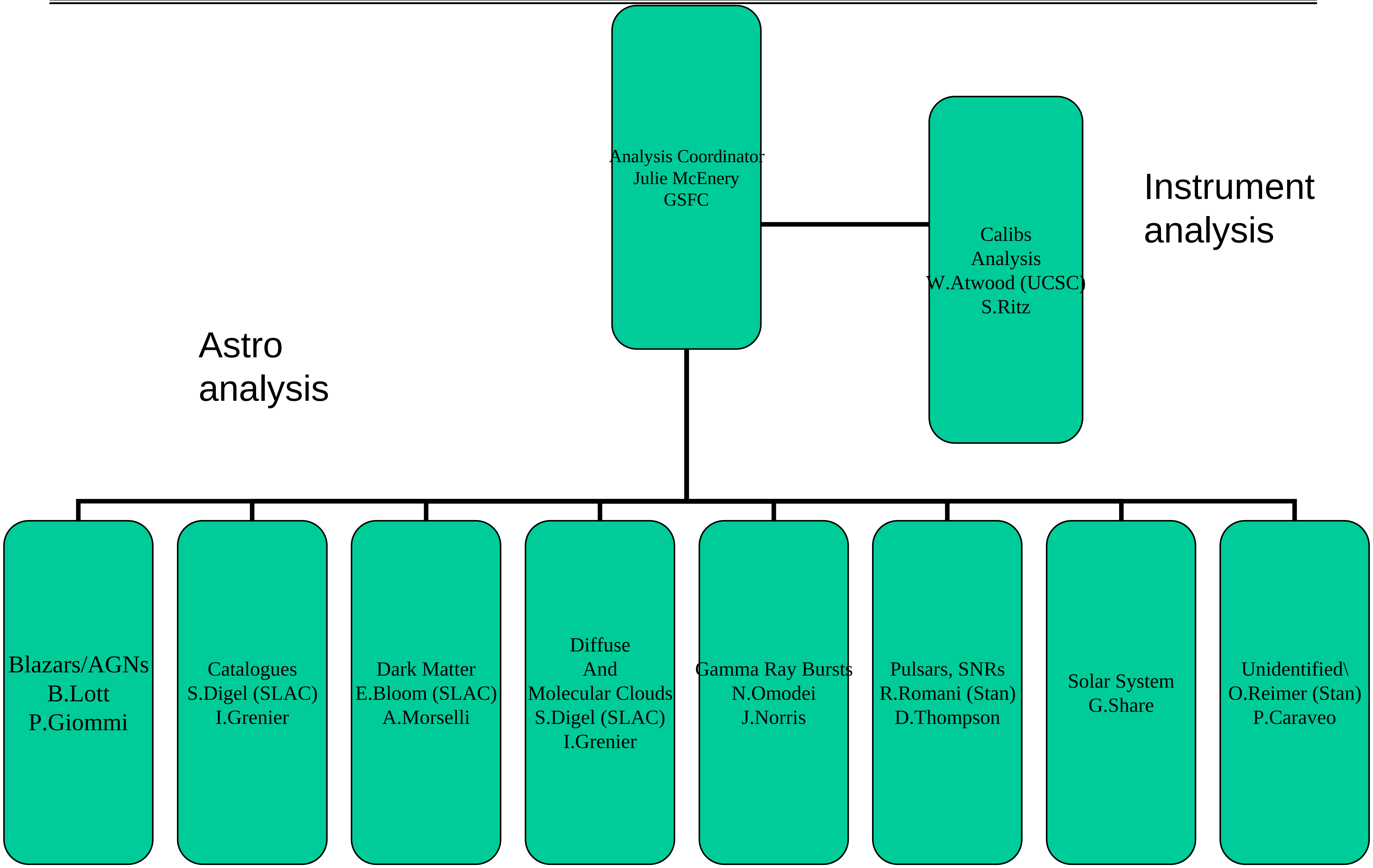
Richard Dubois
Stanford Linear Accelerator Center
richard@slac.stanford.edu

A Glimpse of the Future



APOD 5/31/06

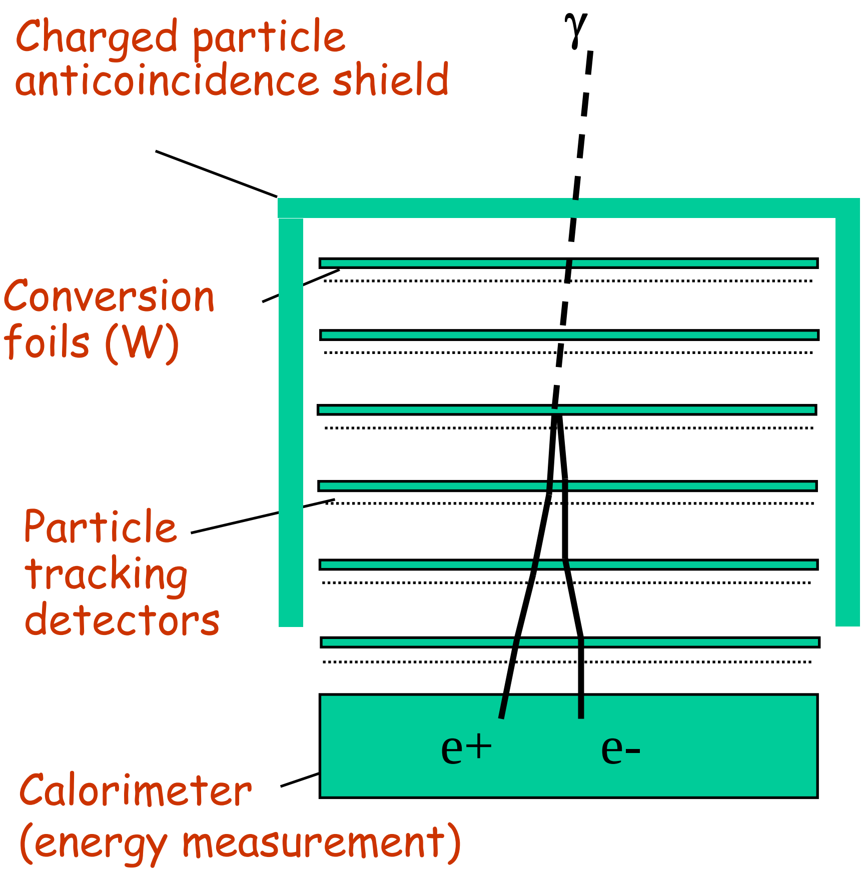
Organizing Science in the Collaboration



GLAST Reconstruction

Anatomy of a "Typical" Event

Pair production is the dominant photon interaction in our energy range



- **Reconstruction Goals:**
  - **Incident Gamma Direction and Energy**
  - **Reject Backgrounds**
- **Incident Gamma converts in the tracker**
  - **In particular, conversion occurs in one of the converter foils – ie at a well defined location**
- **Resulting electron-positron pair range out of tracker (TKR)…**
  - **No magnetic field, tracks are "straight lines"**
  - **Resulting two tracks "point" back to incident Gamma**
- **And into the CsI Calorimeter (CAL)**
  - **Measures total energy of electron-positron pair**
  - **= Gamma energy**
- **Surrounding Anti-Coincidence Detector (ACD) vetoes any wayward charged particles**

GLAST Reconstruction

What makes it challenging…

- **Track Opening Angle ~0**
  - **Resolve**
    ~ 2 * 228 um / 30 mm = ~15 mr

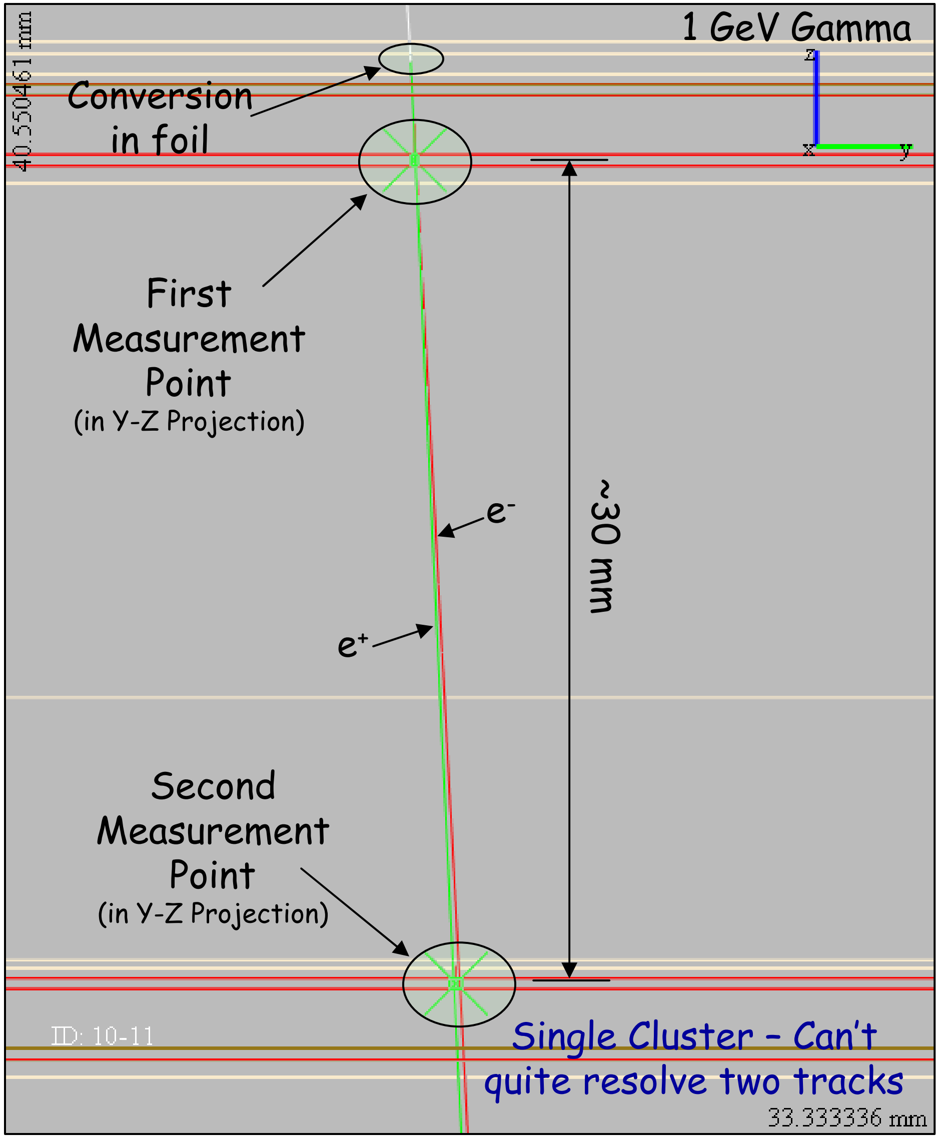
    Strip Pitch

    ~ Tray Spacing

    < ~50 MeV photons to resolve tracks without "help"

- **Looking for "v"s may not be the correct strategy for gamma direction reconstruction**
  - **Well… see next slides…**

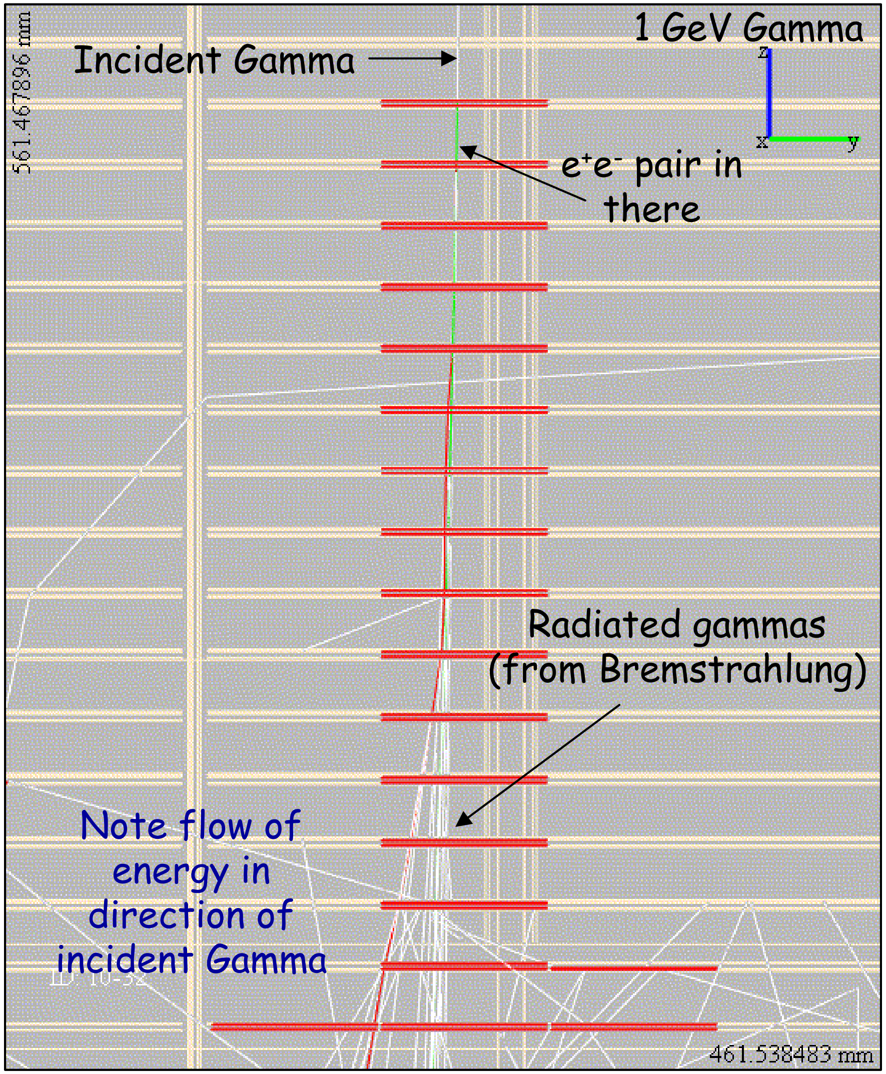
T.Usher



GLAST Reconstruction

What makes it challenging…

- **Tracker has a lot of material**
  - **Actual tracker is ~ .3 rl**
    - **Could live with this…**
  - **Converter foils are ~ 1.1 rl**
    - **Love them: convert gamma**
    - **Hate them: tracking electrons**
  - **Total ~ 1.4 rl**
    - **For particles traversing active area of tracker**
    - **Does not include walls between towers, etc.**
- **Issues to deal with**
  - **Gammas can (and do) convert outside the foils**
  - **e^+e^- pair interact with tracker**
    - **Multiple scatter**
    - **Primary e^+ or e^- can stop in the tracker**
    - **e^+ and e^- radiate energy**
    - **etc.**

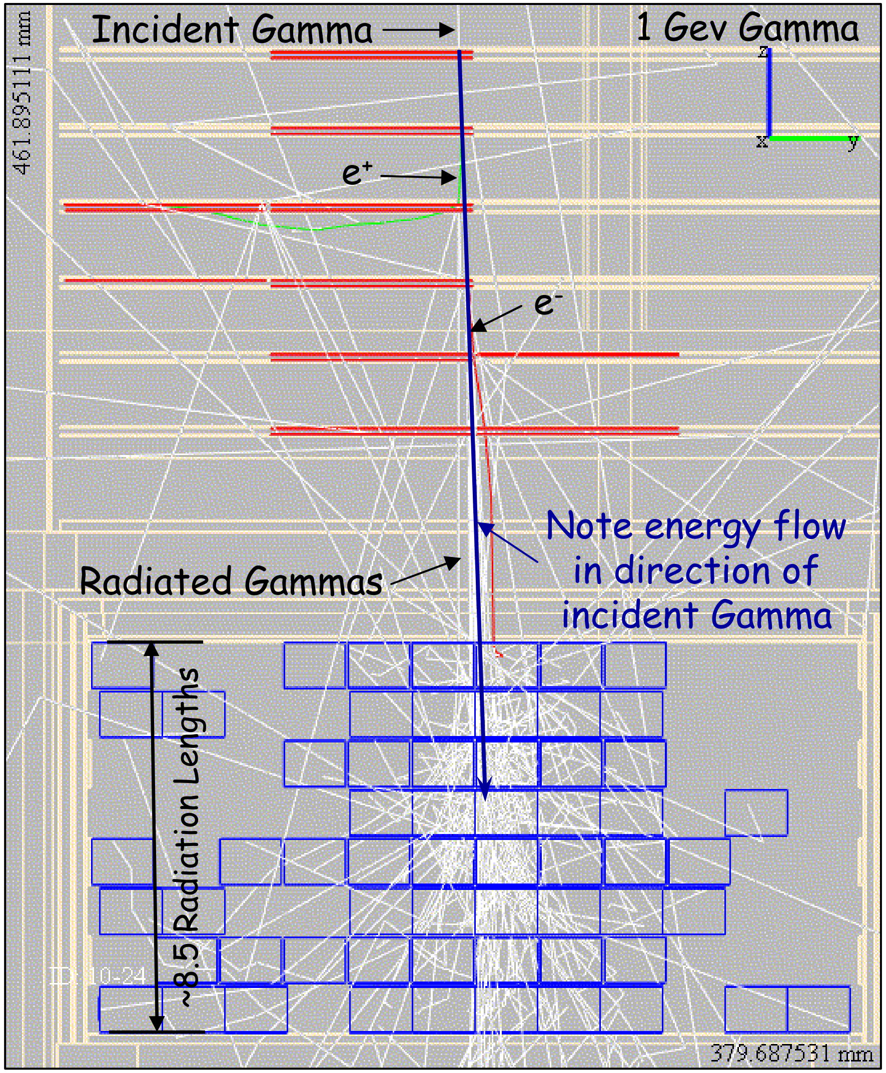


T.Usher

GLAST Reconstruction

What makes it challenging…

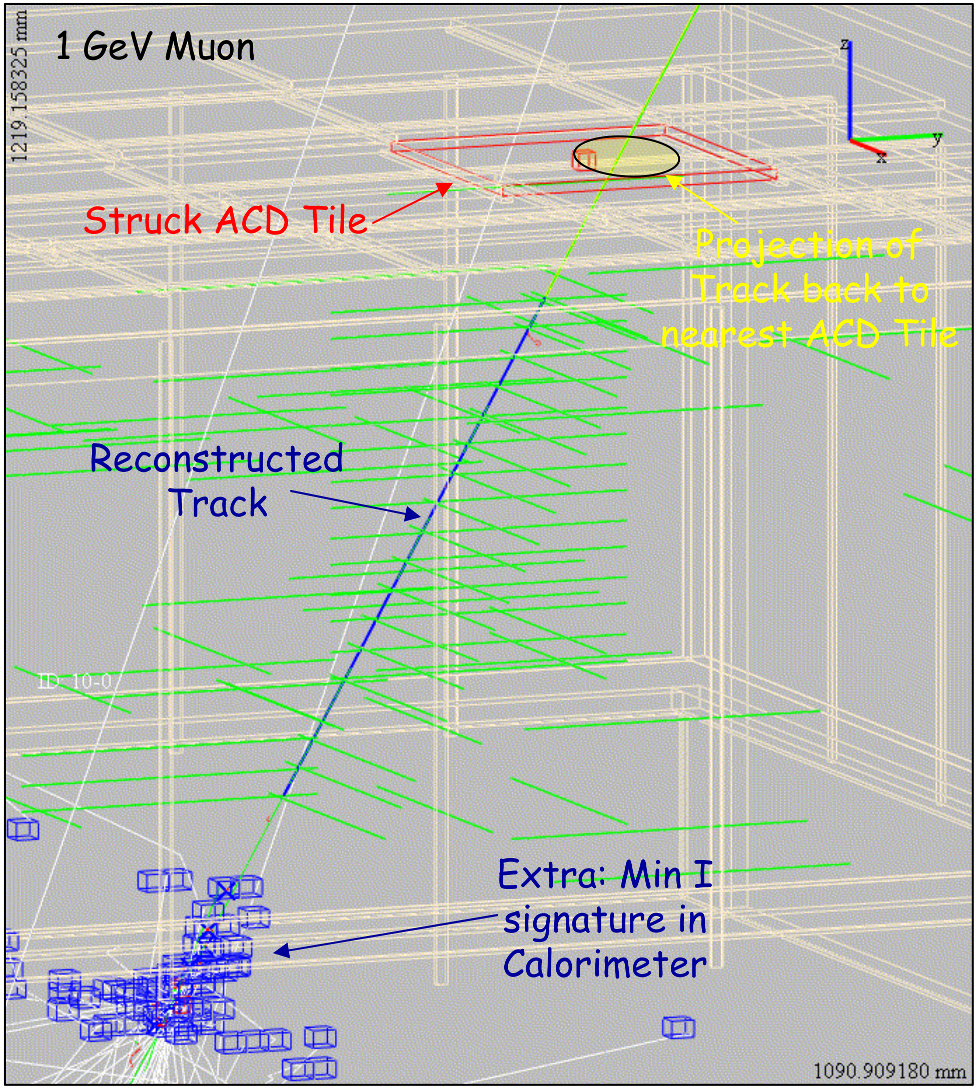
- **Calorimeter Issues**
  - Measure Event Energy – Not Track Energy(ies)
    - Don't have resolution to separate
    - Large fraction of measured energy from Brems
    - Implications for determining gamma direction when you do have two track events…
  - Measure Fraction of Event Energy
    - Energy "loss"
      - in tracker
      - Leaking out of Calorimeter
    - Significant contribution at
      - lower energies (e.g. < 1 GeV)
      - for conversions starting higher in the tracker
    - Must augment total energy determination with contribution from tracker



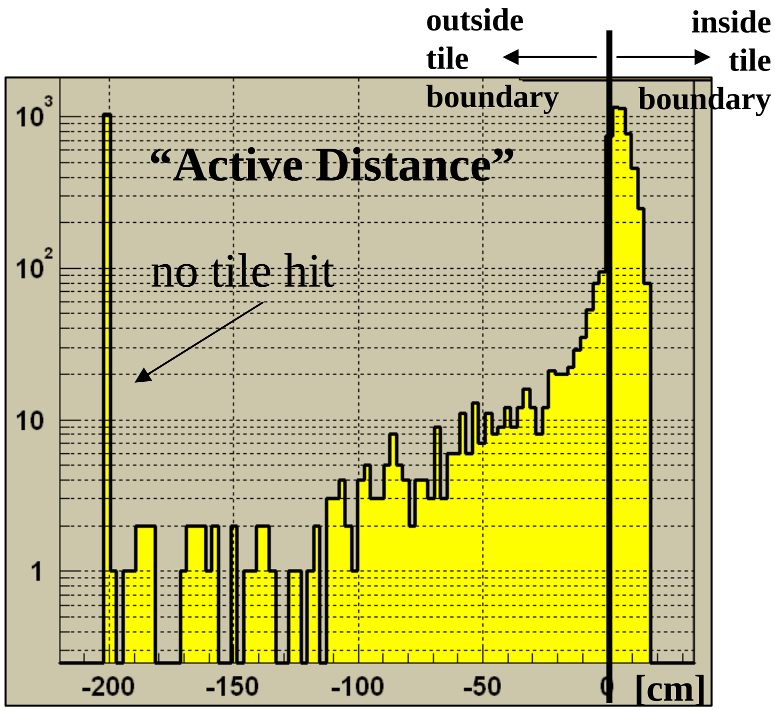
T.Usher

Background Rejection

Example: Charged Particles in Tracker



- Project Track to plane of struck tile
- Calculate distance to nearest edge
- Sign
  Positive if track projection inside the tile
  Negative if track projection outside the tile
- Reject if inside the tile

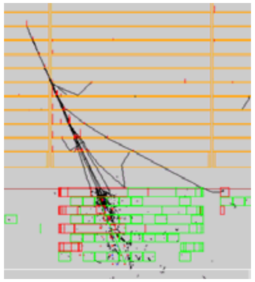


Sim/Recon Toolkit

Package	Description	Provider	Status
ACD, CAL, TKR Recon	Data reconstruction	LAT	90% done In use
ACD, CAL, TKR Sim	Instrument sim	LAT	95% done In use
GEANT4 v8	Particle transport sim	G4 worldwide collaboration	In use
xml	Parameters	World standard	In use
Root 5	C++ object I/O	HEP standard	In use
Gaudi	Code skeleton	CERN standard	In use
doxygen	Code doc tool	World standard	In use
Visual C++/gnu	Development envs	World standards	In use
CMT	Code mgmt tool	HEP standard	In use
ViewCvs	cvs web viewer	World standard	In use
cvs	File version mgmt	World standard	In use

Instrument Simulation and Reconstruction

3 GeV gamma interaction

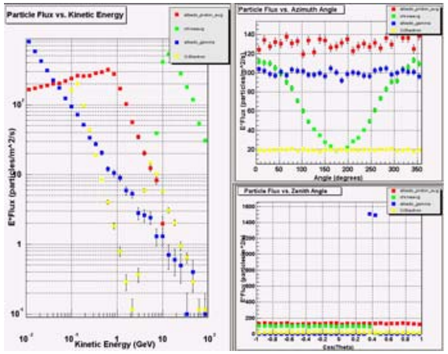


Source Fluxes

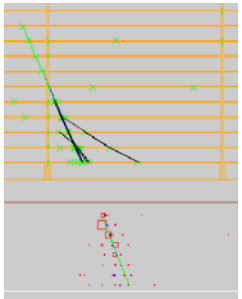
Particle Transport

"Raw" Data

Instrument data



Recon

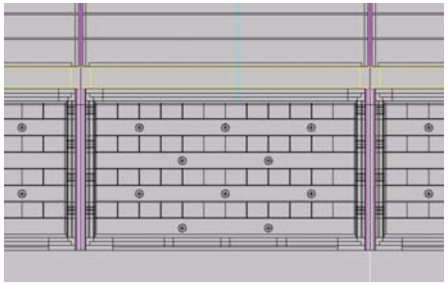


3 GeV gamma recon

Geometry

Background Rejection - Particle ID

Full geometry in xml with C++ interface
G4 discovers instrument from the xml



CAL Detail

Science Tools

- The 'Science Tools' are the high-level analysis tools for astronomy
- The core analysis tools have been defined and developed jointly with the GLAST Science Support Center (NASA/GSFC)
  - NASA staffed the GSSC early with this intent
  - These tools all adhere to the HEASARC FTOOL standards
- To the extent possible we have reused code from existing tools
  - Most notably for pulsar timing, e.g., barycenter arrival time corrections
- For source detection and characterization, the science tools use Instrument Response Functions (PSF, effective area, and energy dispersion as functions of relevant parameters), effectively abstracting the reconstruction and classification process
  - The greatest differences from the formalism for EGRET analysis is that the LAT will almost always be slewing, so that the response functions that apply to any given source also change continuously

Science Tools (2)

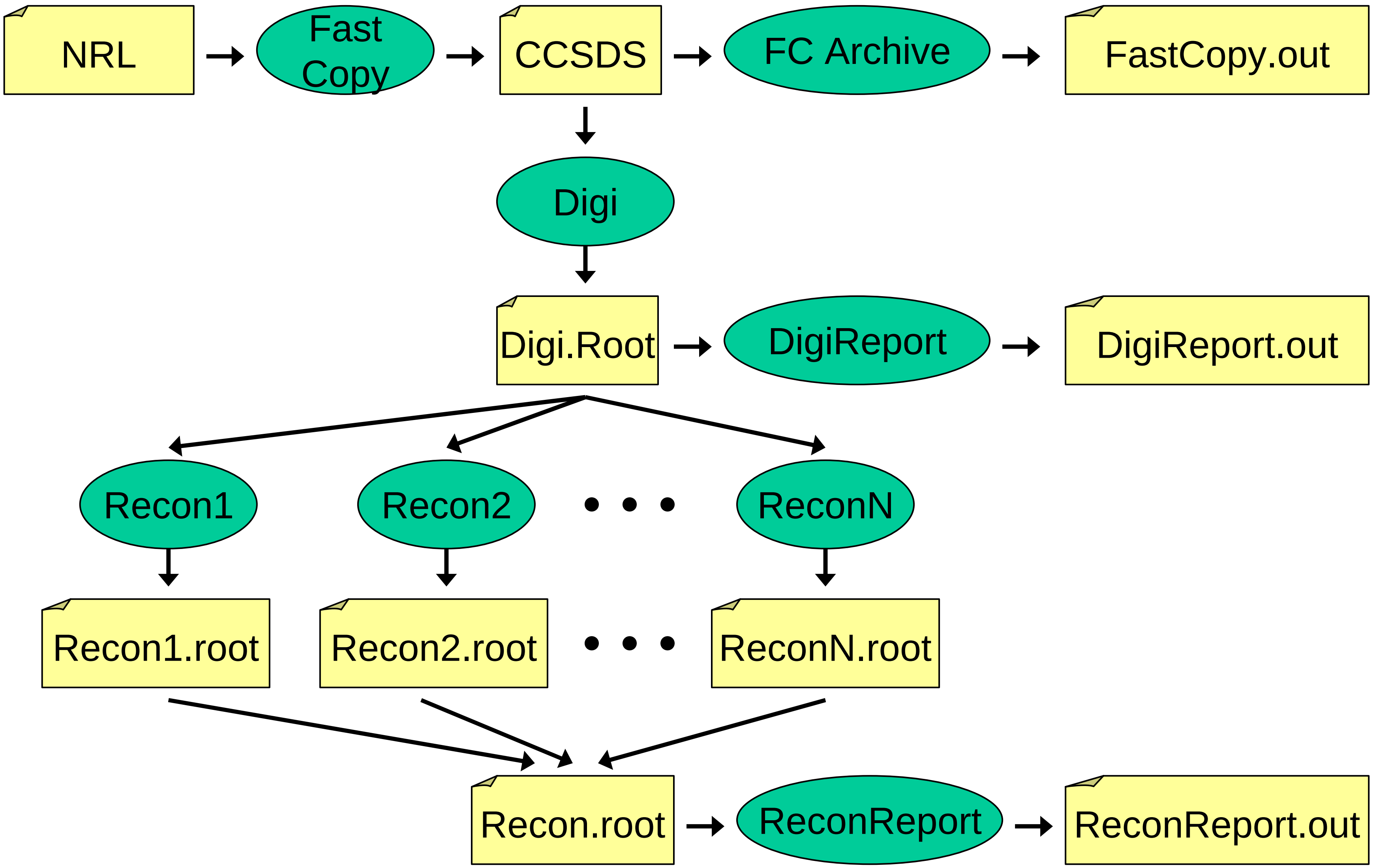
- After a period of definition and review, the tools have been developed incrementally, with the milestones for evaluation
  - Data Challenges (see later) as major milestones and 'Science Tools Checkouts' (3 so far) as intermediate ones
- The core Science Tools are
  - gtlikelihood, gtexpmap, and numerous associated utilities – for defining a model of a region of the sky and fitting it via maximizing the likelihood function
  - gtrspgen, gtbin – for generating response matrices and counts spectra for analysis of GRBs in XSPEC, including jointly with GBM data
  - gtbary, gtpphase, gtpsearch – and associated utilitites for pulsar timing, periodicity tests
  - gtobssim, gtorbsim – fast and flexible observation simulator using the IRFs, and an orbit/attitude simulator.

Automated Pipeline Processing

- What is the pipeline?
  - Envisaged as tool to provide a tree of processing on a given input dataset
  - Handle multiple “tasks” concurrently, eg LAT commissioning, DC2 Monte Carlo runs
  - Full bookkeeping to track what happened
  - Archive all files touched

- Used by whom?
  - Online
    - for sweeping integration data out of the clean room and to tape
    - populate eLogbook
  - SVAC (Science Verification and Calibrations)
    - for doing digi, recon
    - creating reports
    - Preparing for calibrations
  - Generic MC
    - DC2, background runs etc etc
  - ISOC (Instrument Science Operations Center)
    - Flight operations: Level 1 and 2
    - environmental testing, at Spectrum Astro, KSC
    - Data reprocessing

Sample Processing Chain



Web Monitoring of Pipeline Progress

GLAST Pipeline

Version 1.4.2 | Jira | Help

Access control by user

User: richard . Logout

Mode: **Prod** Switch to: [Prod | Dev | Test]

Upload configuration file

Task in question

summary / interleaveDC2-GR-v7r3p21 / mc1step

Processing step in chain

Runs for process: mc1step

NEW Show processing statistics

Filter queries

Run Min [ ] Max [ ] Status: Success

Date Start None End None [Filter] [Clear]

[ ] Show all runs on one page

31,619 items found, displaying 1 to 20.
[First/Prev] **1**, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Run	Status	Submitted	Memory (MB)	CPU (secs)	Job Id	Links (?)
1	Success	2006-02-01 22:38	565	2035	100544	Log : Files : Out : Err
2	Success	2006-02-01 22:45	564	2075	100845	Log : Files : Out : Err
3	Success	2006-02-01 22:45	557	2140	100844	Log : Files : Out : Err
4	Success	2006-02-01 22:45	561	2183	100843	Log : Files : Out : Err
5	Success	2006-02-01 22:45	561	2197	100846	Log : Files : Out : Err
6	Success	2006-02-01 22:45	576	3598	100847	Log : Files : Out : Err
7	Success	2006-02-01 22:45	579	2220	100850	Log : Files : Out : Err
8	Success	2006-02-01 22:45	578	2322	100848	Log : Files : Out : Err
9	Success	2006-02-01 22:45	599	2313	100849	Log : Files : Out : Err
10	Success	2006-02-01 22:45	599	2376	100851	Log : Files : Out : Err
11	Success	2006-02-01 22:45	593	2772	100852	Log : Files : Out : Err
12	Success	2006-02-01 22:45	538	3965	100864	Log : Files : Out : Err
13	Success	2006-02-01 22:45	594	3717	100865	Log : Files : Out : Err
14	Success	2006-02-01 22:45	587	2726	100867	Log : Files : Out : Err
15	Success	2006-02-01 22:45	554	3956	100868	Log : Files : Out : Err

Pipeline 2

Targeted for deployment for beam test

- Build on experience from #1
  - #1 is now robust, but lacking in areas of flexibility
- Revisited requirements:
  - Task scheduling should be more flexible that current linear chain
    - Should support parallel execution of tasks
    - Should allow dependency chain to be more general than the input file requirements
    - Should support parallel sub-tasks, with number of sub-tasks defined at runtime
    - Perhaps support conditions based on external dependencies
  - Should allow for remote submission of jobs
    - Perhaps using GRID batch submission component, or Glast specific batch submission system
    - Will need to generalize current system (e.g. get rid of absolute paths)
  - Support reprocessing of data without redefining task
    - Need way to mark Done task as "ReRunnable"
    - Need to support multiple versions of output files
  - Ability to Prioritize tasks
  - Ability to work with "disk space allocator"
  - Would be nice to set parameters (env vars) in task description
  - Would be nice to be able to pass in parameters in "createJob"
  - Ability to suspend tasks
  - Ability to kill tasks
  - Ability to throttle job submission (ie max number of jobs in queue)
  - Ability to map absolute path names to FTP path names (site specific)
  - Would be nice to remove need for "wrapper scripts"
  - Ability to specify batch options (but portability problems)
- Redesigning database schema now
- Targeting beamtest for production use

Instrument Data Access

GLAST *The Gamma Ray Large Area Space Telescope*

Version 1.1 | Jira | Help

Run Mode: **Prod**

Welcome | **Pruner** | Peeler | History | Admin

Select detailed event data

Select summary ntuple events

Batch data for user: richard

Task type: pruner batch number: 126191

Please edit data and than press the 'Proceed' button

Batch Parameters	Parameter Values
Task Name	interleaveDC2-GR-v7r3p21
E-mail	richard@slac.stanford.edu
Tcut	CTBGAM>0
Min Run Number	
Max Run Number	
Debug mode	false
User Comment	10-day test prune
Batch Options	
Max Filesize [MB]	

Data Challenges

- Ground software is amalgam of HEP instrument software and Astro FTOOLS
- Adopt HEP's "Data Challenges" to create a series of end-to-end studies: create a progression of ever more demanding studies
- Originated by the Mark2 experiment at SLAC while waiting for the SLC accelerator to deliver data
  - Test and oil the data analysis system from simulating the physics through full blown analyses
  - Details of physics and detector performance not revealed to the collaboration until closeout
  - Engage the collaboration and get it thinking science
- ISOC is an integral part of the collaboration
  - Exercise its part and interactions with the rest of the collaboration

Data Challenges: Three Rounds

- **DC1. Modest goals. Contains most essential features of a data challenge.**
  - 1 simulated day all-sky survey simulation
  - find GRBs
  - recognize simple hardware problem(s)
  - a few physics surprises
  - Exercise all the components
- **DC2, kickoff Mar 1. More ambitious goals. Encourage further development, based on lessons from DC1. Two simulated months.**
  - DC1 +
    - Much more data
    - Backgrounds included
    - More realistic GRBs
    - Pulsars, variable AGNs
    - More and more elaborate surprises
- **DC3, in CY07. Support for flight science production.**

DC Components

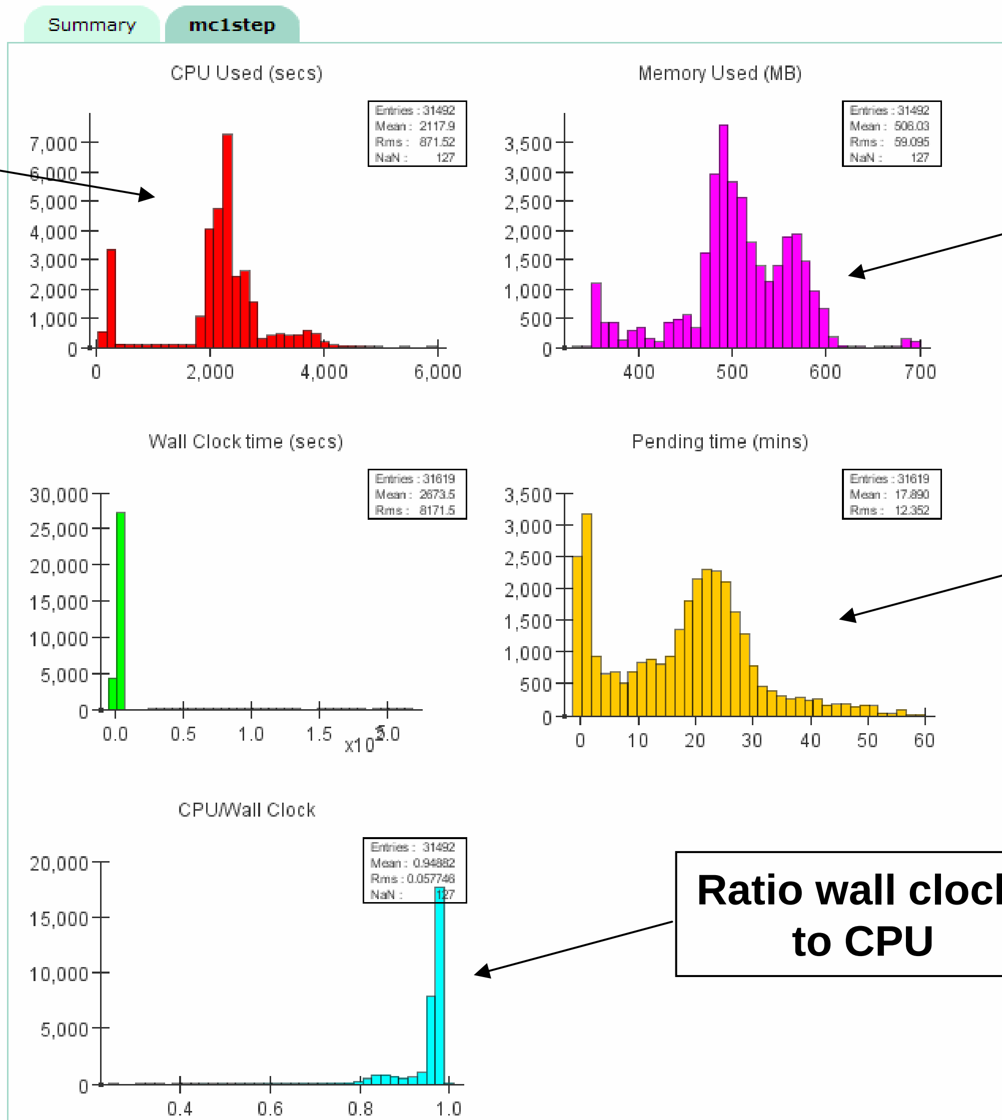
- Focal point for many threads – "end to end system test"
  - Orbit, rocking, celestial coordinates, pointing history
  - Plausible model of the sky
  - Background rejection and event selection
  - Instrument Response Functions
  - Data formats for input to high level tools
  - Use of Science Tools
  - Generation of datasets
  - Populate and exercise data servers at SSC & LAT
  - Code distribution on windows and linux
- Involve new users from across the collaboration
- Teamwork!

Preparations for DC2

- Full background analysis this time!
  - Tremendous collaboration effort to reduce the backgrounds to Science Requirements levels
  - Revision of background model – x4 higher than DC1 estimate
  - Detailed skymodel
    - Flaring objects; pulsars; joint GBM data(!); etc etc
  - Mechanically a huge change from DC1
    - Have to simulate backgrounds 10^3 x signal
    - 100,000 CPU-hrs to simulate 1 day of background: 5 billion events
    - Machinery to randomly interleave that day 55 times, while simulating full rate deadtime effects
    - High-stress test of processing pipeline
      - ~400 CPUs running simultaneously for a week for the backgrounds runs
      - ~200,000 batch jobs total for DC2
    - Many scaling problems fixed

Monitoring Pipeline Throughput

Pipeline Statistics



CPU time

Memory

"Wait" time for jobs

Ratio wall clock to CPU

Monitoring Disk Farm via SCS Tools

sulky35.slac.stanford.edu Overview

This host is up and running.

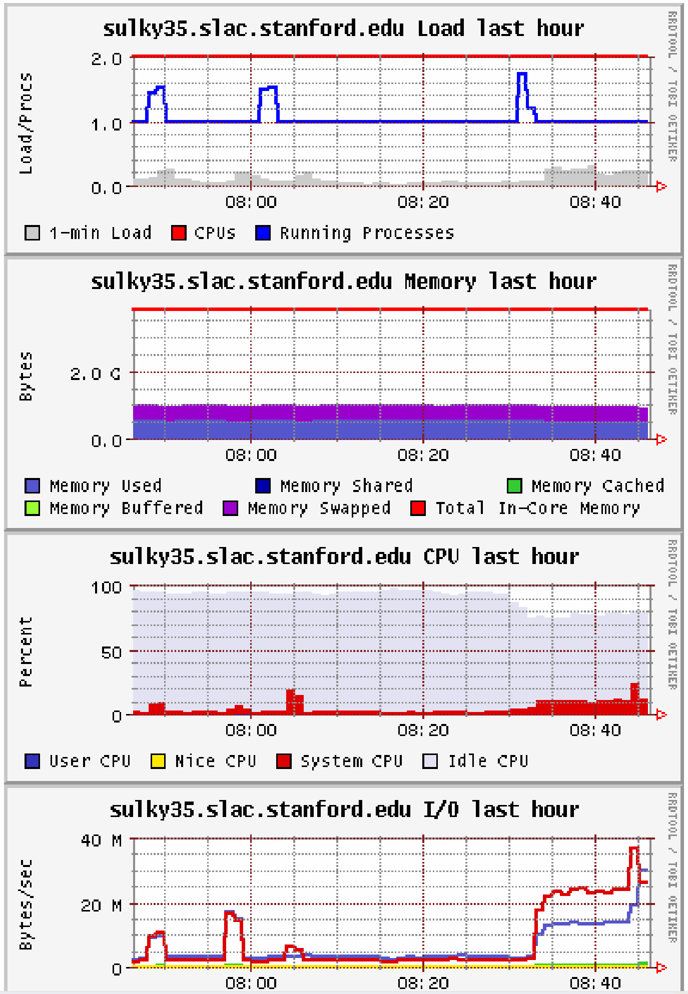
Time and String Metrics

boottime	Sun, 5 Feb 2006 14:25:42 -0800
gexec	OFF
last_reported	0 days, 0:00:07
machine_type	sun4u
os_name	SunOS
os_release	5.9
uptime	6 days, 18:20:32

Constant Metrics

cpu_num	2 CPUs
cpu_speed	1503 MHz
mem_total	4010424 KB
swap_total	11805840 KB

Gmetrics



Documentation: User Workbook

GLAST Data Challenge II

SITE MAP

go

Home: DC2 | Movie: DC2 Sky | Getting Ready | DC2 Code Releases | DC2 Data | DC2 Data Access | Event Viewer | Feedback Wiki & Help | People & Planning | Science Tools

Welcome to Data Challenge II !

Kickoff March 1-3, 2006

Key Links: Kickoff Agenda | DC2 Workshop | DC2 Users Forum | DC2 Analysis Results

Closeout Meeting: May 31 - June 2, 2006
Closeout Agenda

Second in a sequence of 3 data challenges, DC2 will include extensive analysis of 55 days worth of realistic simulated data. Relative to DC1, many aspects of the simulated data have been improved

DC2's "Simulated Gamma-ray Sky" featured as APOD's picture of the day: May 31, 2006.

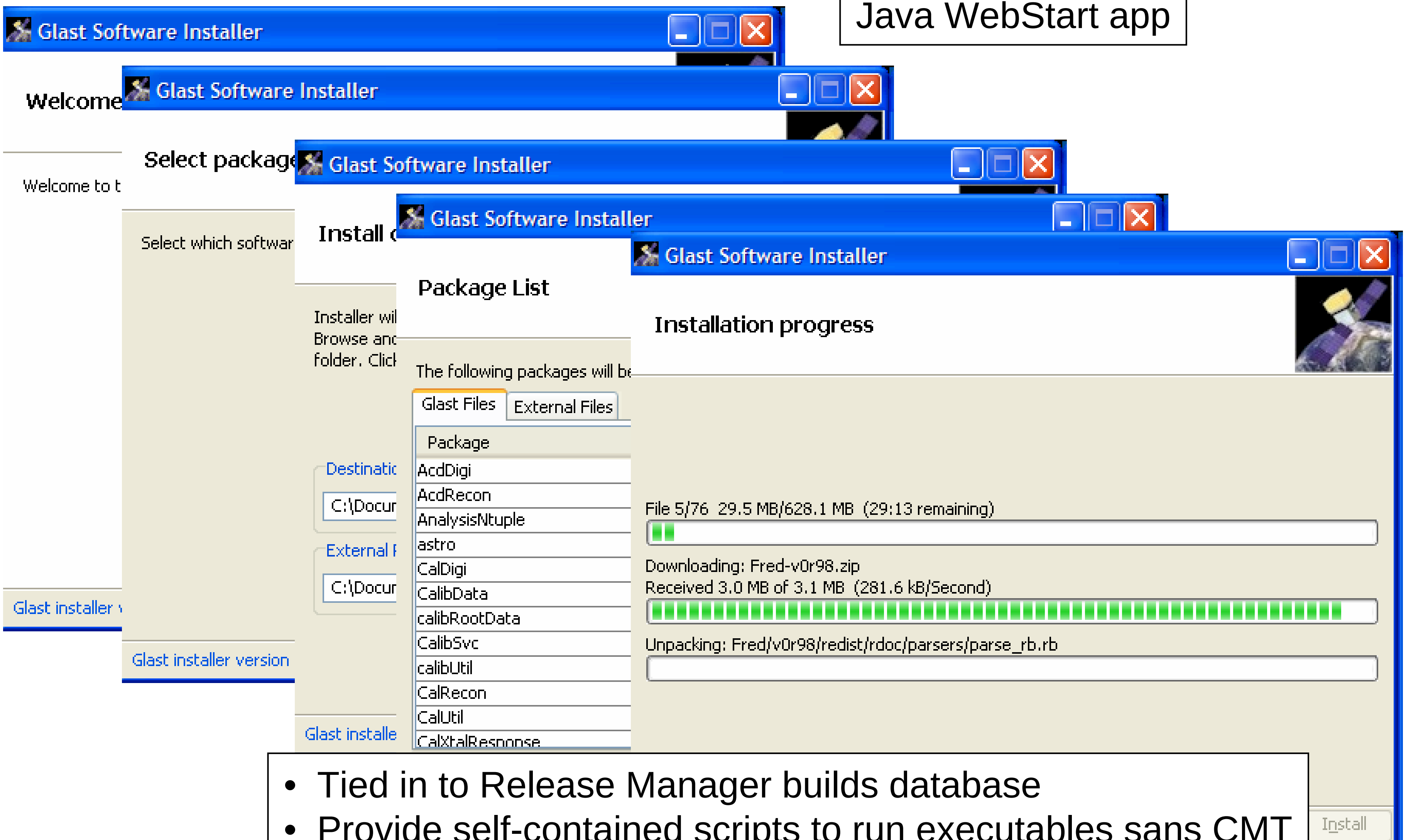
- **Full and realistic detector simulation,** i.e., realistic in that simulations are imperfect, including things like dead strips, deadtime, etc.
- **Full reconstruction chain** from low level electronics quantities, to high level event information.
- **Updated and detailed background model,** including all the things that are not gamma-rays from celestial sources; also includes orbit variations.
- **Rich description of the gamma-ray sky,** including variable sources, pulsing sources and Gamma-Ray Bursts (GRBs).
- **Revamped event classification and background rejection analysis,** i.e., updated instrument performance plots.

Follow on lead from SLD, BABAR, but …

- work with Tech Writer
  - skilled at extracting information from us wackos
  - worries about layout, organization
  - can write good
- we're struggling with apparent conflict of web navigation vs "printed book". Pursuing the former.

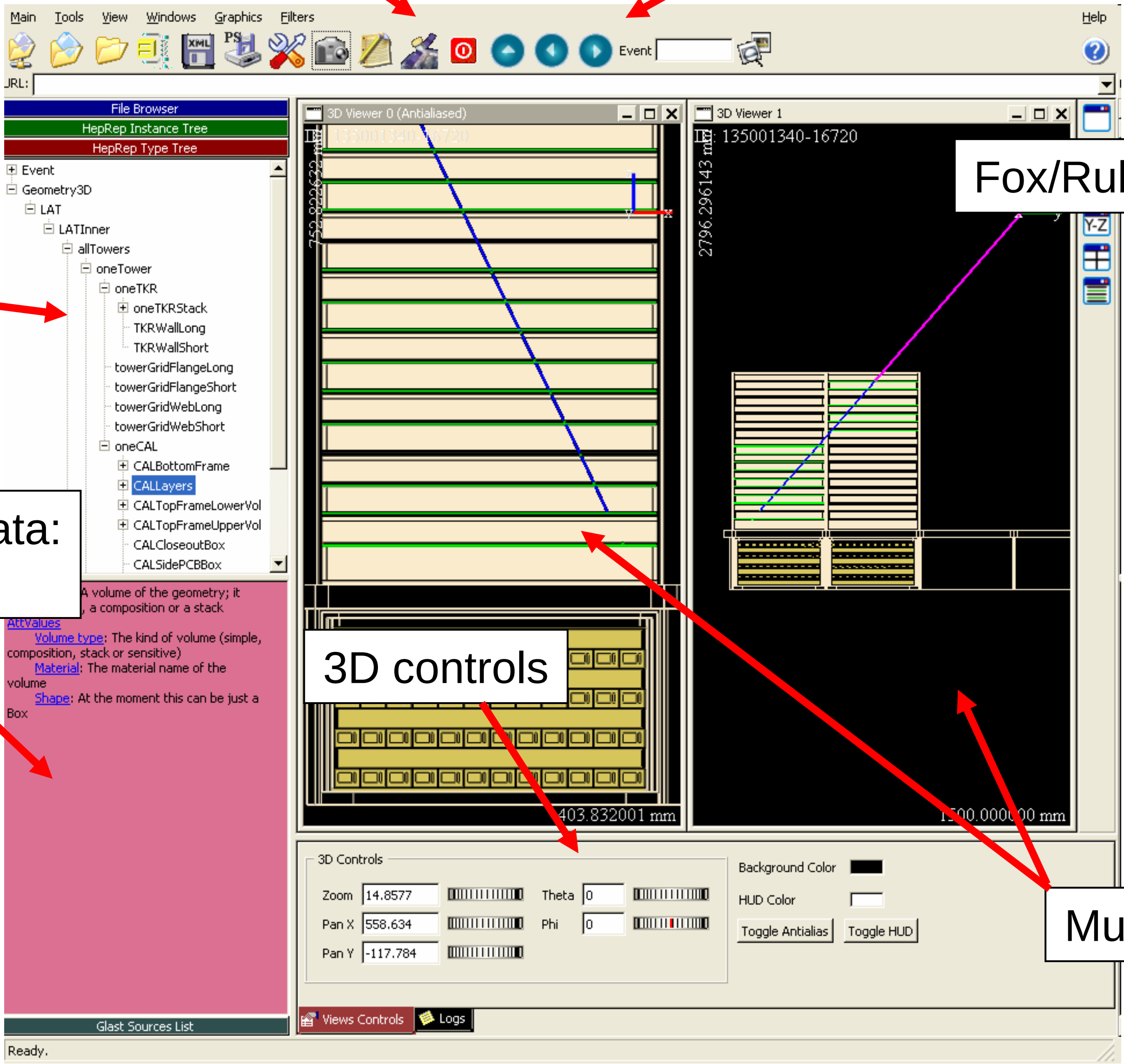
Code Distribution

Java WebStart app



- Tied in to Release Manager builds database
- Provide self-contained scripts to run executables sans CMT

FRED – Event Display



DC2 Kickoff Meeting 1-3 March

112 registered attendees!

France: 13
Italy: 17
US: 71
Japan: 5
Sweden: 2
Germany: 4 (GBM)

Data Challenge II

Logo by Stefano Ciprini

http://www-glast.slac.stanford.edu/software/DataChallenges/DC2/JuneCloseout/

Closeout Meeting: 31 May – 2 June
DC2 Coordinated by Julie McEnery, GSFC (I'm liberally swiping slides from her closeout talk!)

DC2 Goals, requirements and purpose

- 55 days of LAT data provide a deeper view of the high energy gamma-ray sky than has previously been achieved.
  - Results from previous gamma-ray missions provide, at best, an incomplete guide to the DC2 sky.
  - Part of the challenge of DC2 will be to figure out what was included in the sky model.
  - DC2 data has a fairly realistic level of detail which will support a wide variety of both science and instrument performance studies.
  - Exercise the science tools – but don't feel restricted to them
  - Improve the documentation and analysis software from user feedback.

Gamma-ray sources in the DC2 Milky Way

- With the exception of Pulsars, which were based on a population model and a lot of research and fiddling, we included only likely examples each source class
  - Typically associated with an already-known source (sorry Olaf & Patrizia) without attempting a pop. synthesis
- 'Other 3EG' means that we included all non-spurious sources from the 3rd EGRET catalog (Hartman et al. 1999) even if we did not have a specific counterpart in mind

γ-rays (A+B)*

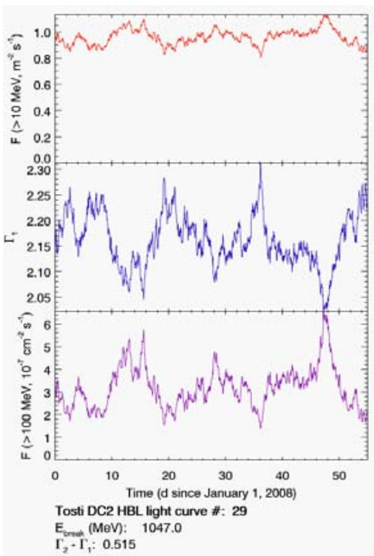
Milky Way itself (1)	1,704,807
Pulsars (414)	140,596
Plerions (7)	9780
SNR (11)	22,592
XRB (5)	1491
OB associations (4)	295
Small molecular clouds (40)	1741
Dark matter (~2)	5158
'Other 3EG' (120)	112,386
Sun (1 flare)	4669
Moon (1)	10,523

*Out of 3,340,146

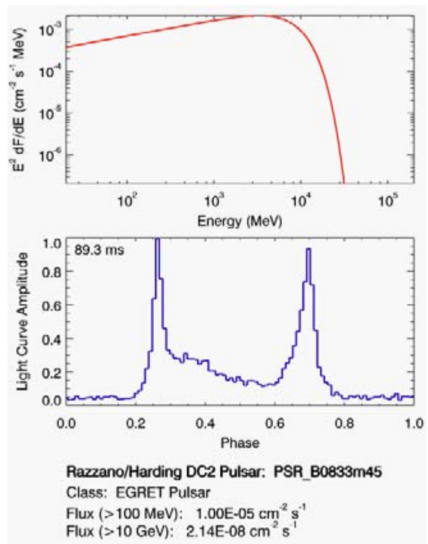
S.Digel

Examples of Variable Sky

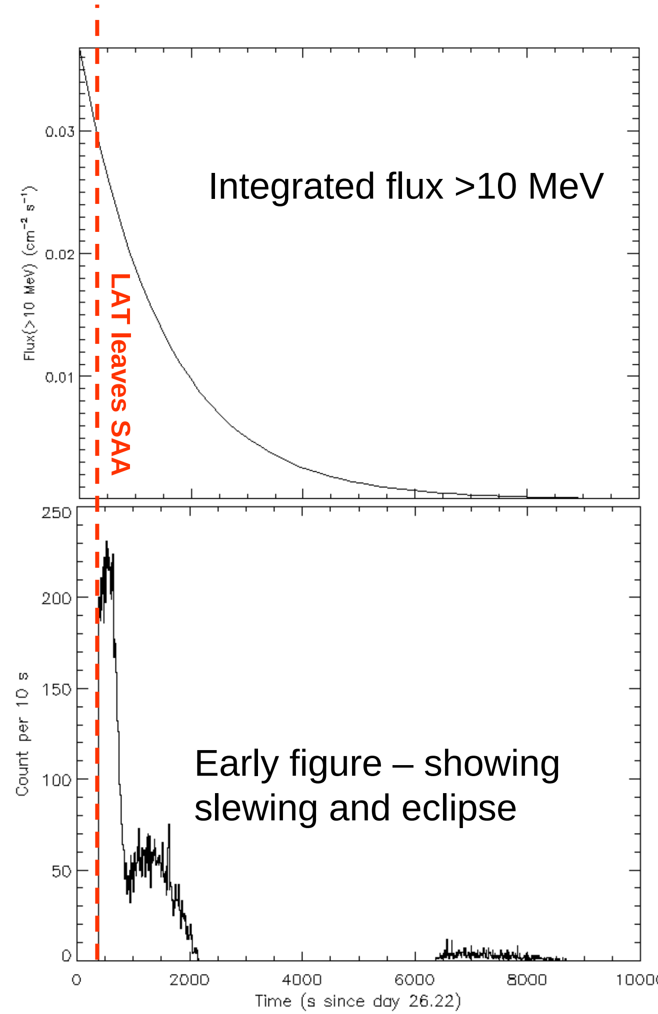
AGN: Mk 421



Pulsar

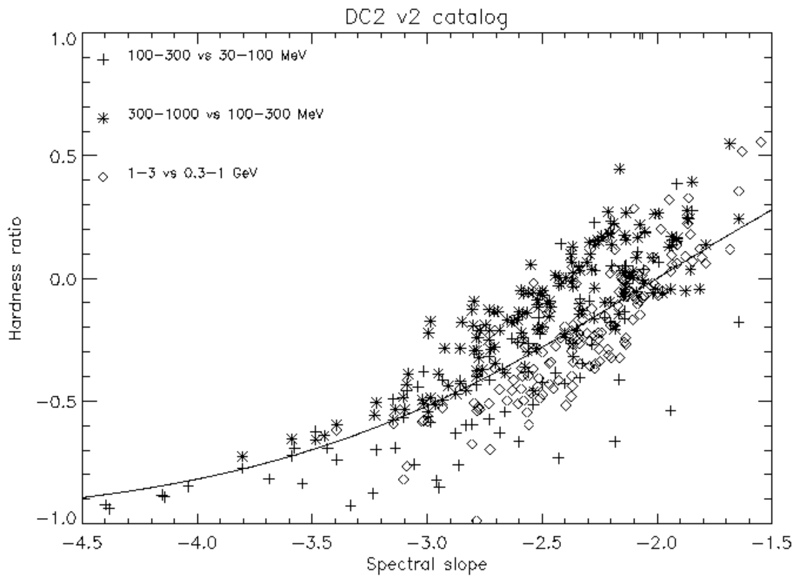


Sun



Produce LAT point source catalog

- Requirement: Spectral index and flux (with associated uncertainties), location with 68% and 95% confidence ranges, flux in discrete energy bands.
- Goal: Variability index, flux history, peak flux, measure of whether a source is extended.

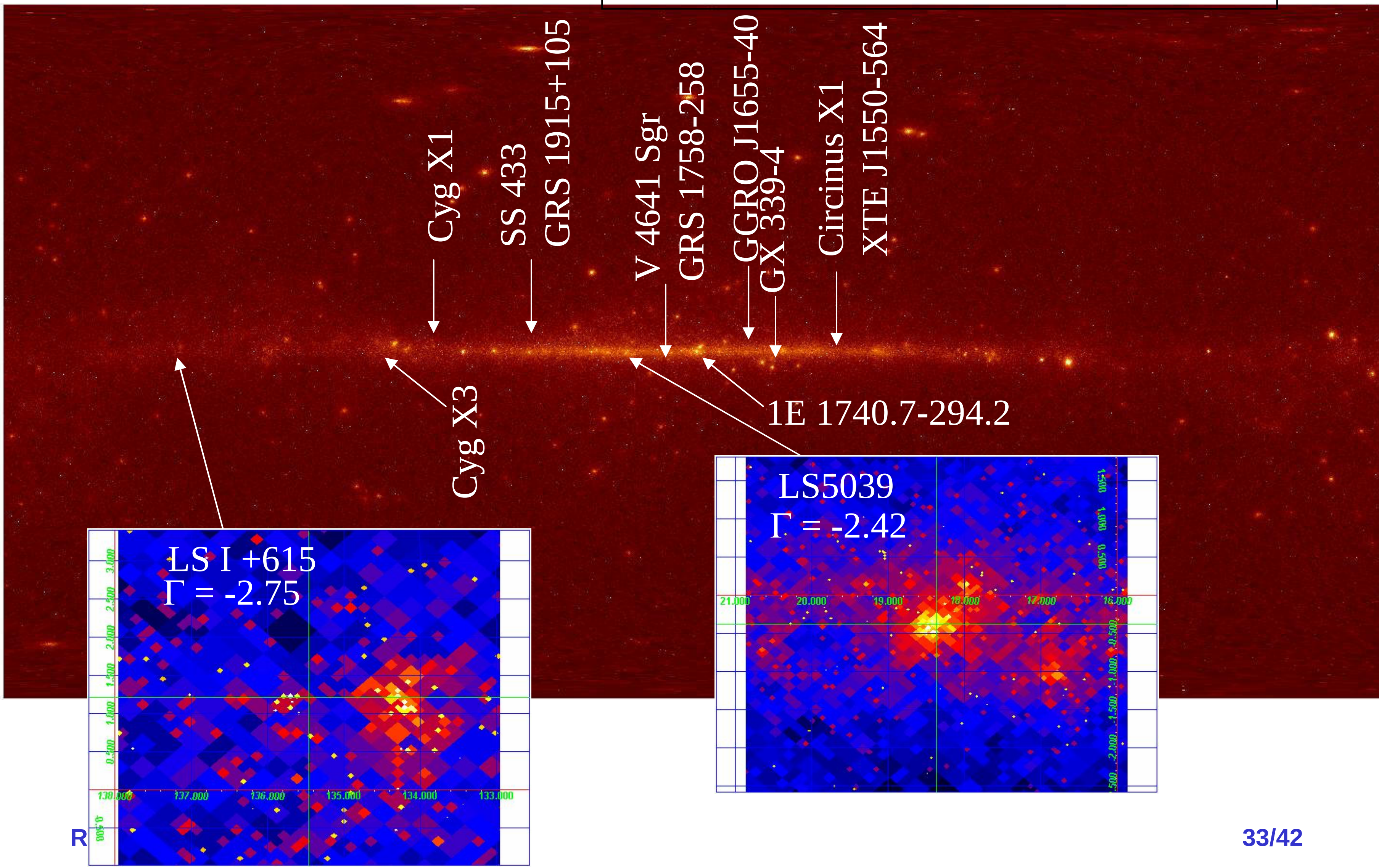


The catalog analysis and results proved to be an extremely important part of DC2. It provided a starting point for a large fraction of the more detailed source analysis and was a reference for people doing population/source detection type studies.

There was a somewhat higher rate of false detections than would have been expected (~10%), this needs to be understood.

LS 5039, LS I +615 & Friends Micro-quasar candidates in DC2

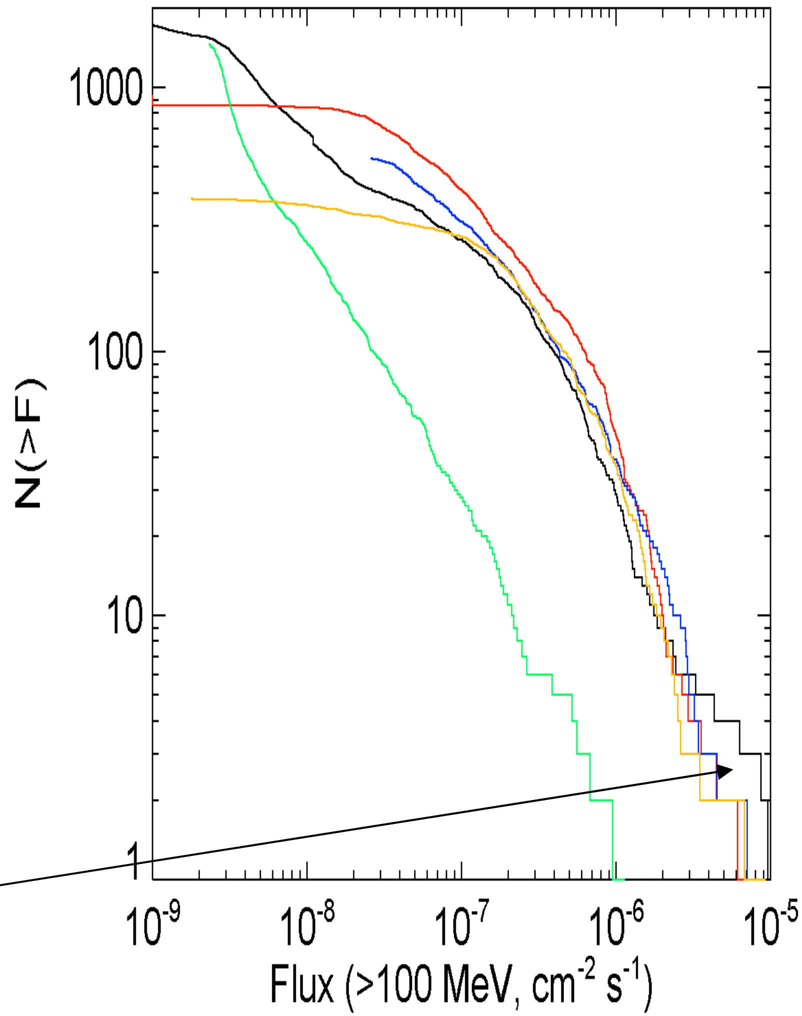
Toby's HEALPIX map + Saclay sources



Develop and test source detection algorithms

- Requirement: That these algorithms are tested and compared with one another in a systematic way using the DC2 data.
  - Many source detection methods developed – Stephens, Tosti, Burnett, Casandjian, Ballet, Romeo/Cillis
  - Compared with one another by Seth Digel

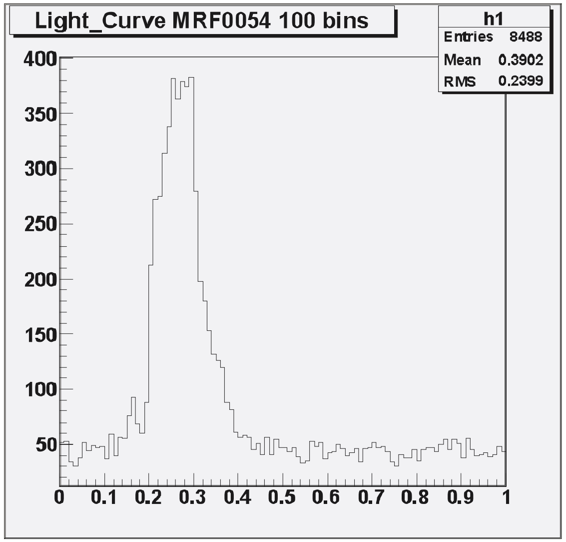
What's going on here?

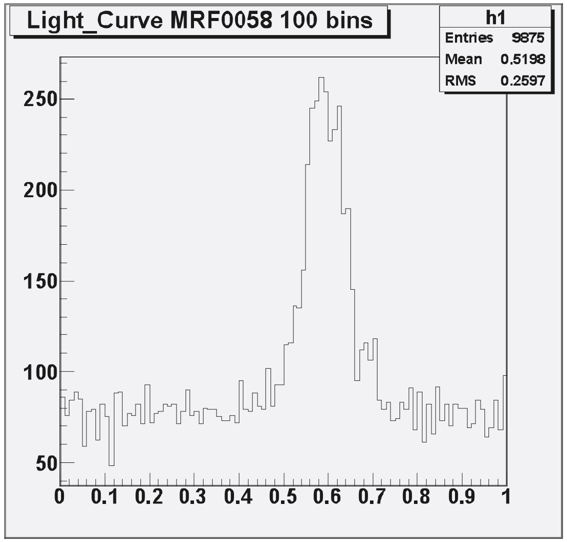


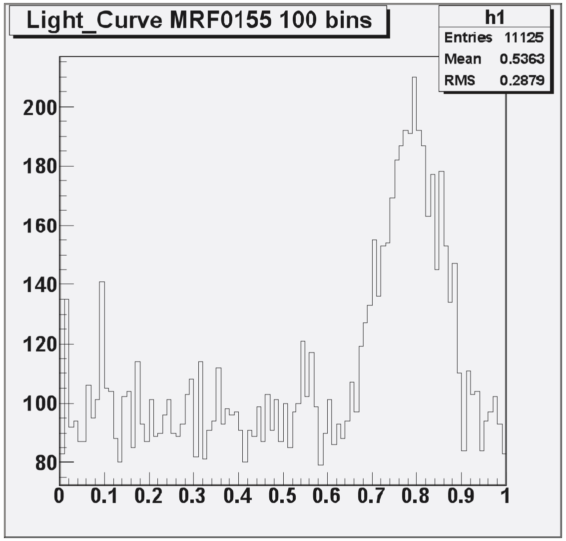
Pulsars

- Goal: blind periodicity searches on candidate DC2 pulsars
  - Use time differences to measure power
  - Look for frequency at peak power

Marcus Ziegler – lightcurves of pulsars without radio data.








```
Epoch_MET =  220838550
F0=  5.885928323969
F1= -1.306230        e-012
F2=  1.0             e-021
```

```
Epoch_MET =  220838550
F0 =  3.91691474178
F1 = -1.936137       e-013
F2 =  6.0            e-022
```

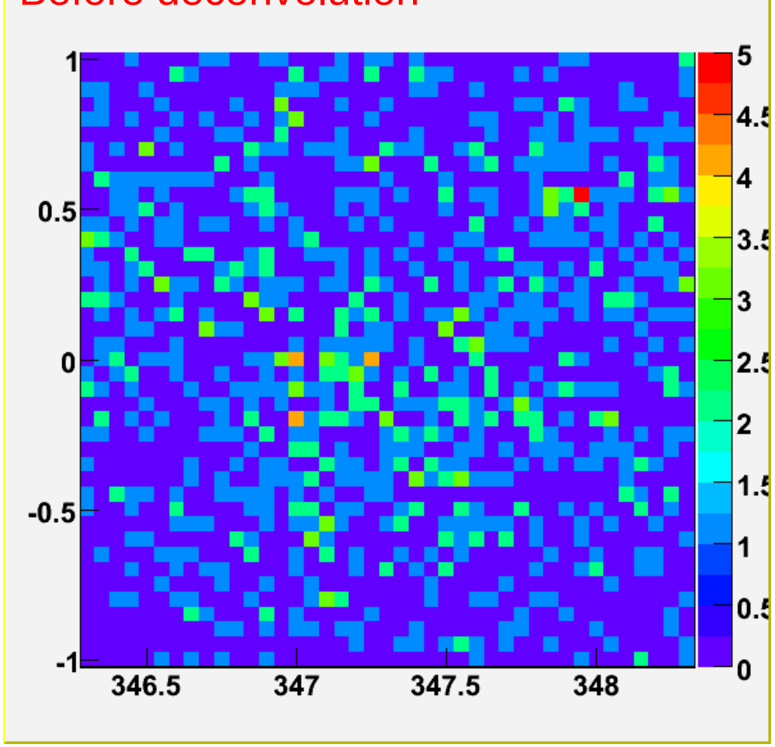
```
Epoch_MET =  220838550
F0 =  3.766282209980
F1 = -3.677283       e-013
F2 = -3.3            e-021
```

Phased light curves for radio quiet pulsars

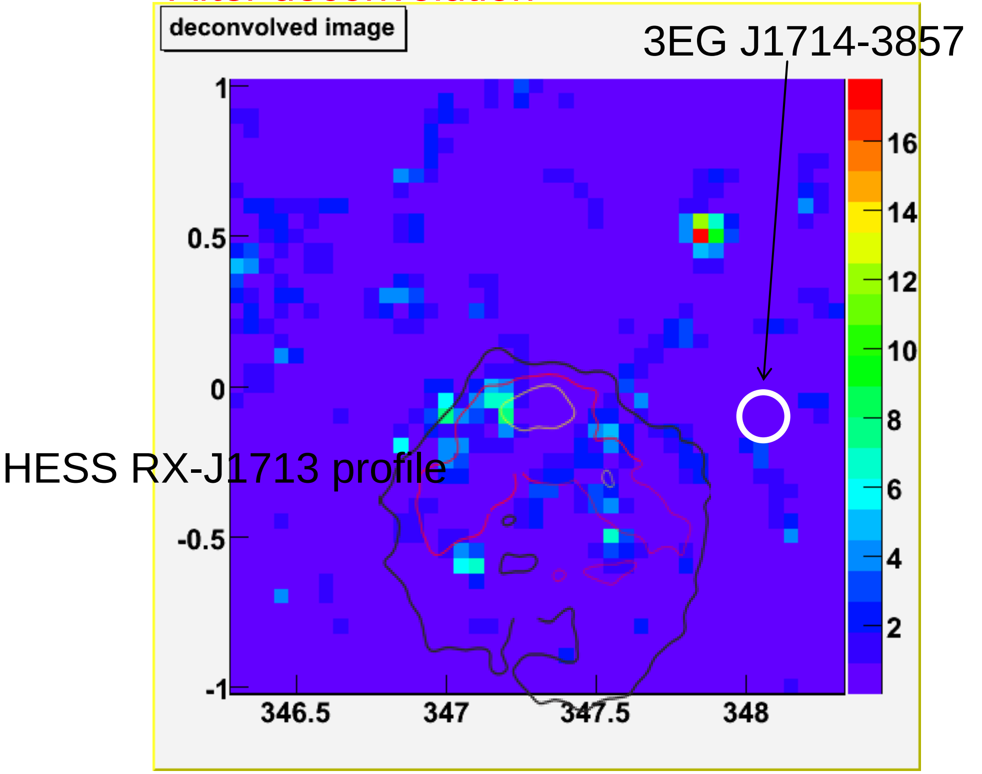
Extended sources

– Hiro studying how to improve images by deconvoluting with the PSF
  - Can we use event by event measured errors?

Before deconvolution



After deconvolution



Hiro Tajima

Variable sources

– Requirement: Produce lightcurves for at least 20 bright sources (from the data release plan, these are the sources we will release high level data from in year 1)

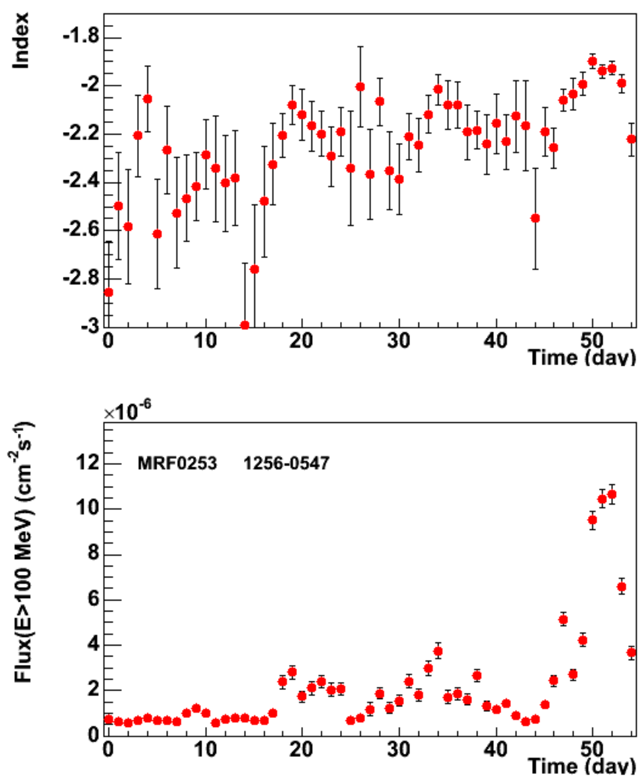
– Goal: look at lightcurves for many more sources

Light curves for sources from the 1st Year Data Release Plan - DC2 - SLAC Confluence - Microsoft Internet Explorer

https://confluence.slac.stanford.edu/display/DC2/Light+curves+for+sources+from+the+1st+Year+Data+Release+Plan

Source name	other name	v1 Catalog name	DRMNGB	MINUIT	simple estimate
0208-512	3EGJ0210-5055	MRF0294	light curve, file	light curve, file	light curve
PKS 0528+134	3EGJ0530+1323	MRF0194	light curve, file	light curve, file	light curve
0827+243	3EGJ0829+2413	MRF0264	light curve, file	light curve, file	light curve
Mrk421	3EGJ1104+3809	MRF0404	light curve, file	light curve, file	light curve
3C 273	3EGJ1229+0210	MRF0409	light curve, file	light curve, file	light curve
3C 279	3EGJ1255-0549	MRF0253	light curve, file	light curve, file	light curve
1406-076	3EGJ1409-0745	MRF0224	light curve, file	light curve, file	light curve
PKS1622-297	3EGJ1625-2955	MRF0362	light curve, file	light curve, file	light curve
1633+383	3EGJ1635+3813	MRF0258	light curve, file	light curve, file	light curve
1730-130	3EGJ1733-1313	MRF0020	light curve, file		light curve
3C 454.3	3EGJ2254+1601	MRF0293	light curve, file	light curve, file	light curve
LSI +61 303	3EGJ0241+6103	MRF0044	light curve, file	light curve, file	light curve
Mrk501		MRF0257	light curve, file	light curve, file	light curve
W Com	3EG1222+2841	MRF0234	light curve, file	light curve, file	light curve
1ES 1959+650		MRF0012	light curve, file	light curve, file	light curve
1ES 2344+514		MRF0351	light curve, file	light curve, file	light curve
H 1426+428		MRF0240	light curve, file	light curve, file	light curve
PKS2155-304		MRF0330	light curve, file	light curve, file	light curve

http://www.cenbg.in2p3.fr/ftp/astropart/glast/DC2/light_curves/sources/lc_5_0_1256-05

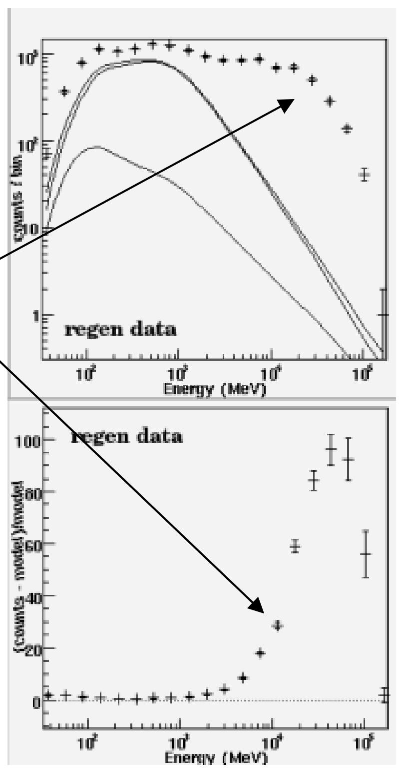


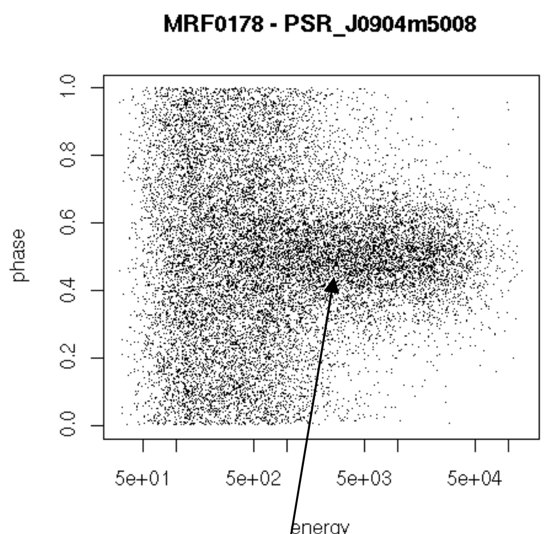
By Benoit Lott

Spectral Studies

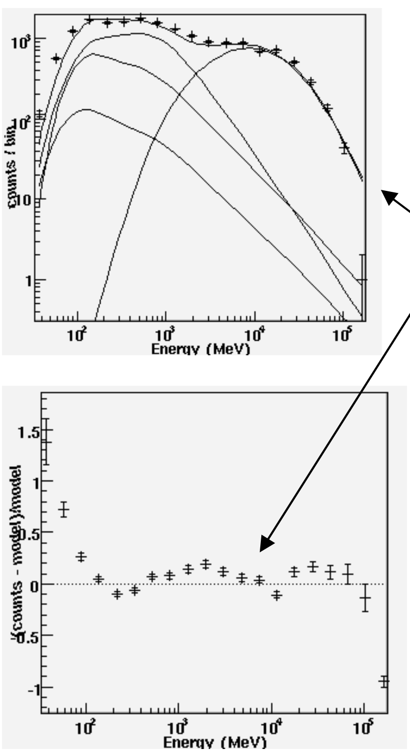
- **Riccardo Rando found a source that appeared to consist of two components, a pulsed hard component and a soft, steady component.**

Power-law point source + background model is a very poor fit to the data



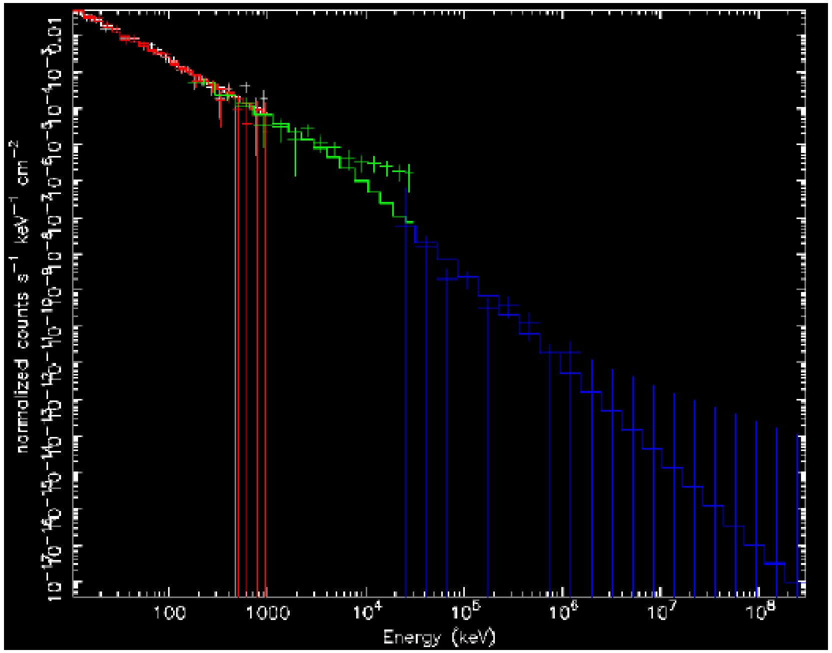


Phase vs energy plot shows that the pulsed emission dominates above 1 GeV



Refit with a composite source consisting of a power-law and a log normal component

Gamma-ray bursts



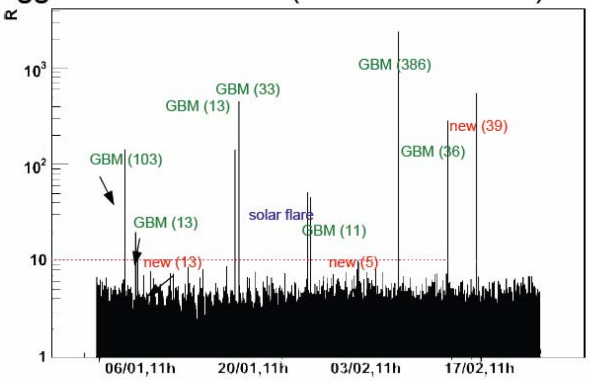
This was one of the “rejected” fits due to the strange spectrum. The cause is likely to be because this GRB was simulated with an additional “hard” extended component lasting for 400s.

GRB08015885 – Nukri Komin

132 generated in 4π
64 bursts seen in GBM
25 in LAT; 16 with > 4 γ

Blind GRB Search

- spikes in LAT trigger rate R = 10/Δt (rate for 10 events)
- 7 GBM bursts
- 3 new bursts
- 1 solar flare

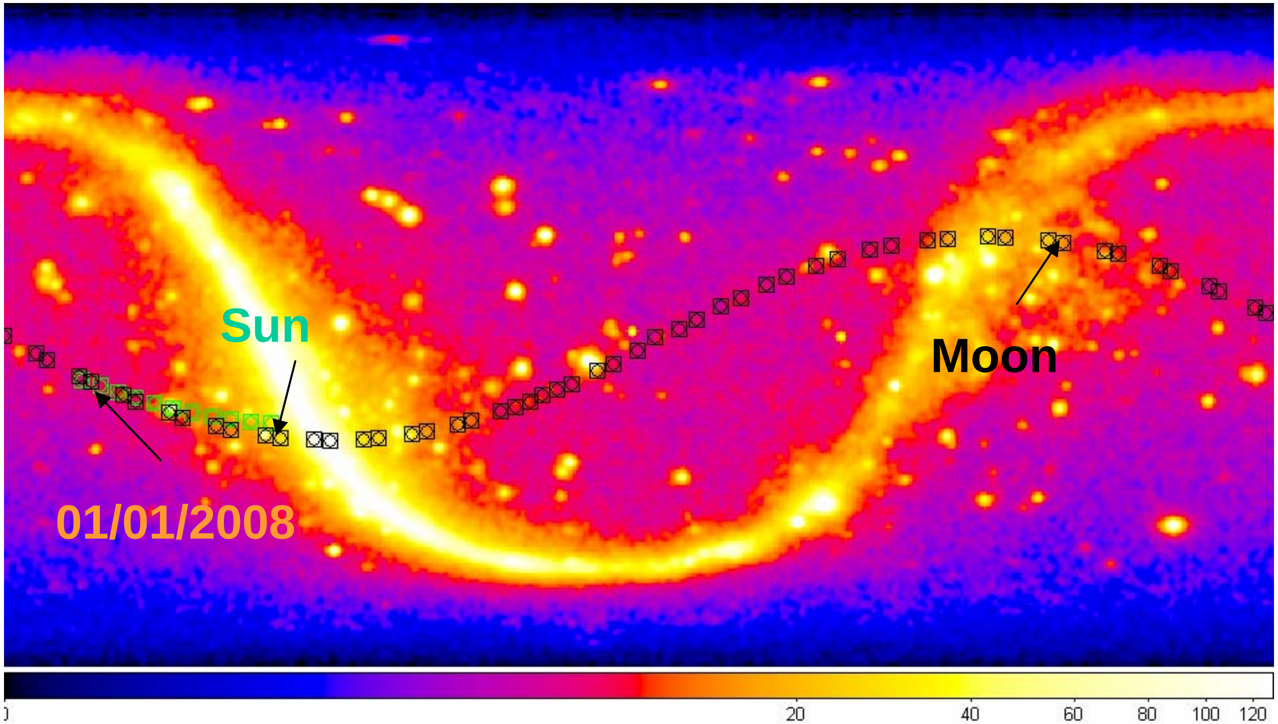



```
root [0] .L GRBsearch.C+
root [1] GRBsearchMulti(
end with ';', '@':abort > "LAT_allsky*.fits.root"
end with ';', '@':abort > 10,10)
```

DC2 close out, GSFC, 31/05/06

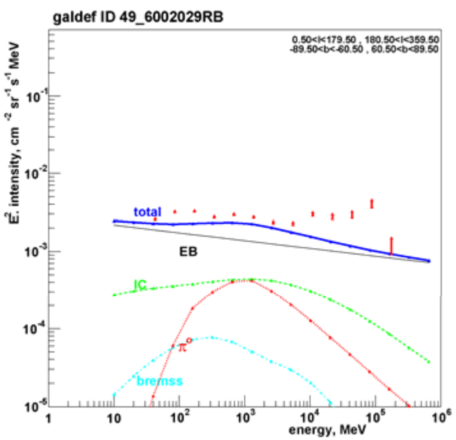
Other sources

- Requirement: Identify at least one source that is not a pulsar, AGN or GRB (there are some that can be identified from the gamma-ray data)
- Moon (Tosti, Rando)
- Sun (Tosti, Chiang)

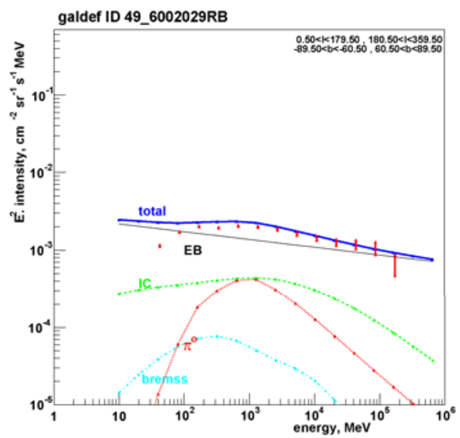


Diffuse sources

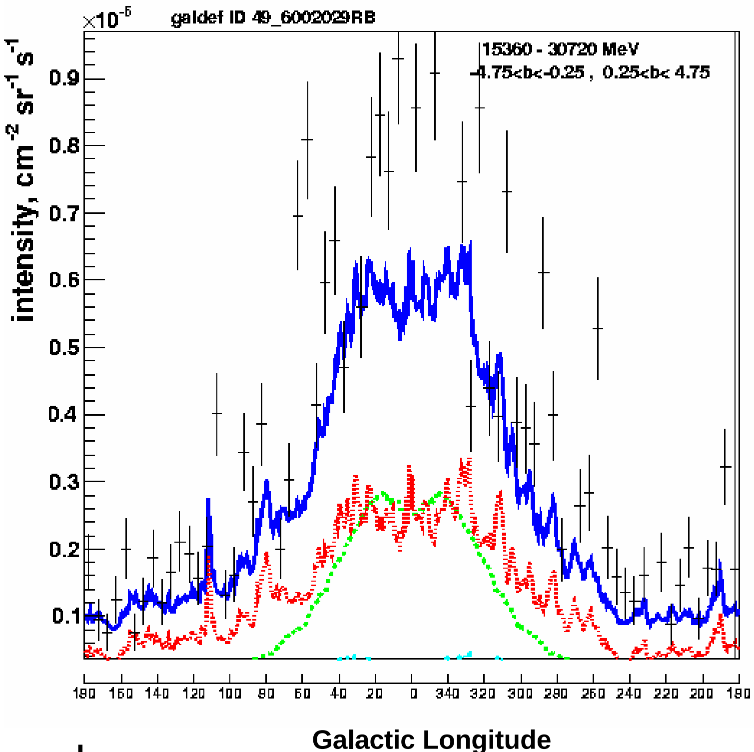
- Goal: Study flux, spectra and spatial distribution of the galactic diffuse and compare with the diffuse model provided for source analysis.
- Studied by Jean Marc Casanjian, Andy Strong and Larry Wai



Full sim



Fast sim
No resid bkg



Diffuse is background to non-line dark matter searches

DC2 and Beyond

- The DC2 sky is probably the best rendition to date of the gamma ray sky
  - LATers took up the challenge and didn't just look for the obvious
- And… we now have a great dataset for future development!
  - 55 days of simulated downlink to practise with
  - Simulate downlink frequency
  - Test Data Monitoring
  - Develop Quicklook
- And then "DC3"