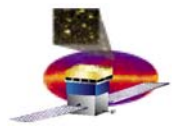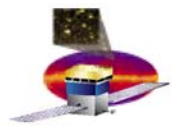# LAT PDR Review: SAS
Higher-Level Processing and the Science Support Center

**S. W. Digel (Stanford, recently ex GSFC/USRA)**

Outline

- **What comes after event reconstruction and classification?**

- **Considerations for astronomical analysis of LAT data**

- **Some implications for high-level tools**
  - **Model fitting**
  - **Interstellar emission model**

- **Infrastructure of the analysis environment**
  - **Science databases**
  - **Science tools**
  - **Observation simulators**

- **GLAST Science Support Center**

- **Project Data Management Plan**
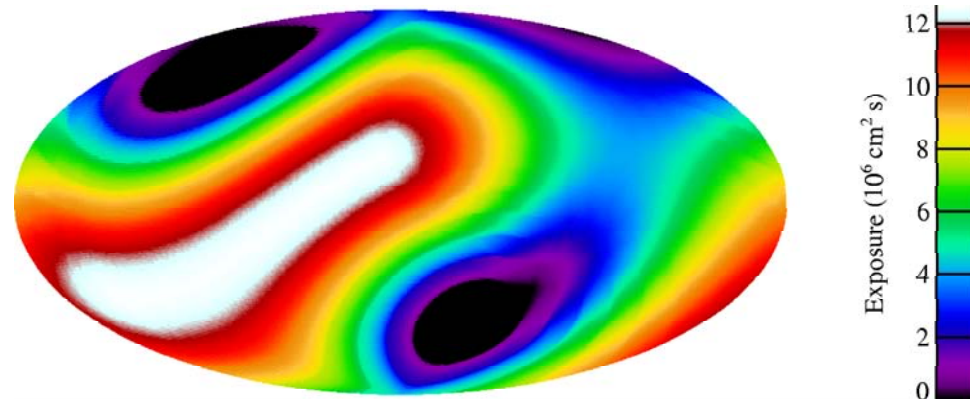
- **LAT Team-SSC Coordination**
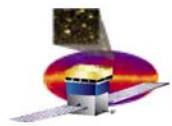
What Happens Post-Level 1?

- **Study celestial sources of high-energy gamma rays:**
    - **Measure positions, fluxes, spectra, time profiles for transient or periodic sources, and extended emission (Milky Way and extragalactic)**

Photons detected in one orbit ($\sim 2 \times 10^4$)
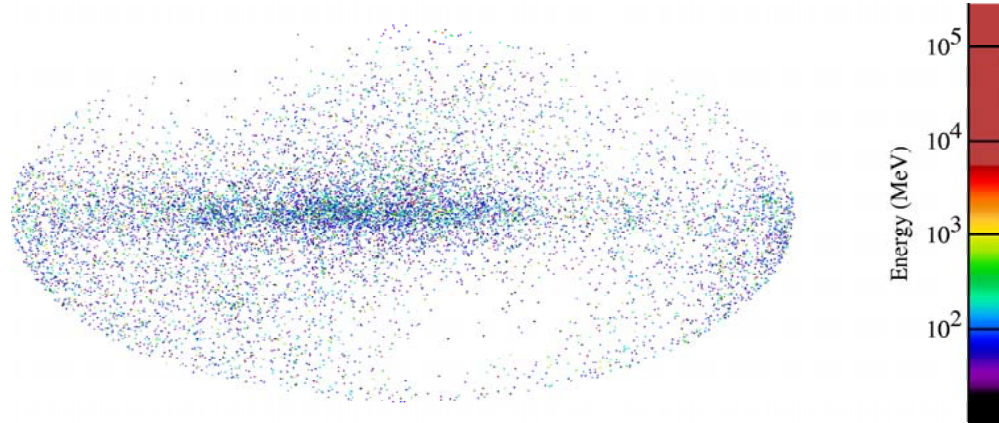
Exposure, one orbit, no rocking, 28.5° inclination, SAA, phony A_{eff}, one energy
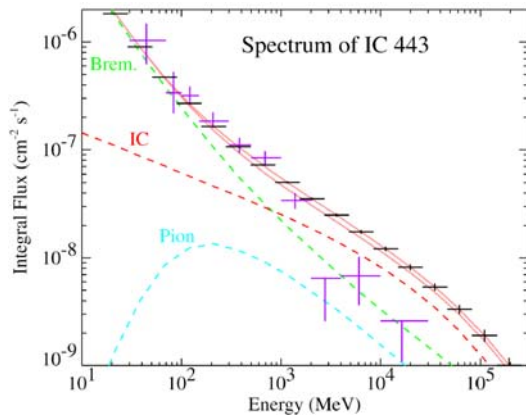
What Happens Post-Level 1? (2)

Source Detection/Characterization
Simulated high-latitude LAT field with isotropic distribution of AGN, one-year sky survey exposure. Contours of source detection significance are shown.
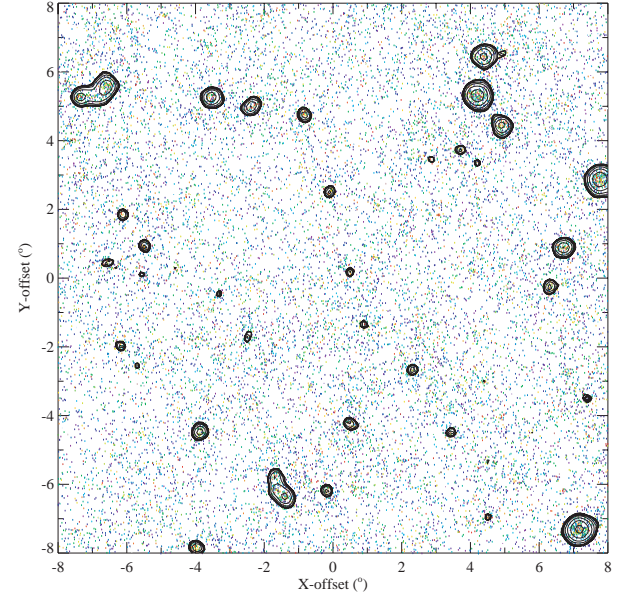
Spectral Measurements
Simulated LAT spectrum of supernova remnant IC 443 with EGRET observation (Esposito et al.) and model components (Baring et al.)
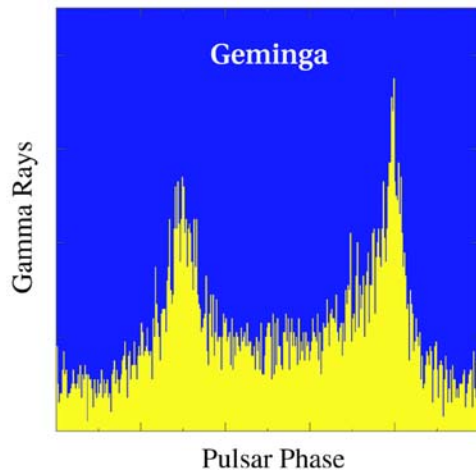
Pulsar Phase Folding
EGRET >100 MeV data for Geminga, period 237 ms (D. Thompson, GSFC)

Time Profiles of AGN
Simulated LAT measurement of flux and spectral evolution of 3C 279 during a flare detected by EGRET (C. Dermer, NRL)

Considerations for Science Tools

- **LAT has very large FOV and extremely broad energy range** $\longrightarrow$

  - **Careful calibration of LAT response over broad range of angles and energies is needed for useful results**

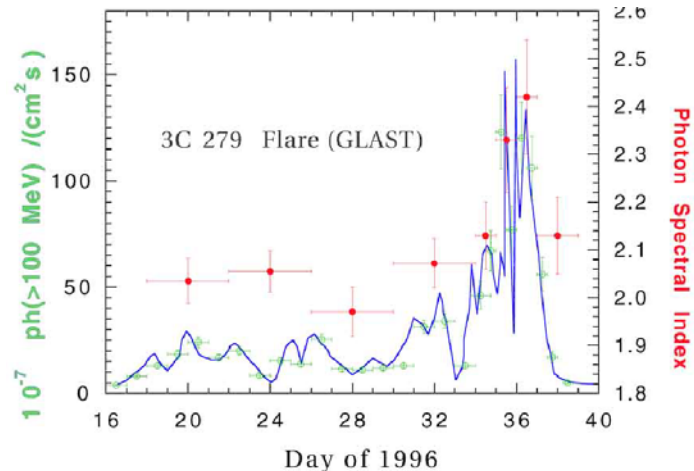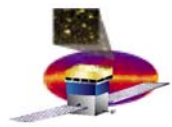- **PSF is really bad (for a non γ-ray astronomical instrument); celestial fluxes are low, and charged particle background is intense** $\Longrightarrow$

  - **High-level analysis of LAT data is fundamentally model fitting**
  - **Interstellar emission model is essential**

- **To maximize efficiency (and operate as a sky monitor), LAT will generally be slewing continously** $\Longrightarrow$

  - **Data for a region of the sky will be far from time ordered and will not even be represented by the same set of IRFs.**

- **LAT will have guest observers** $\longrightarrow$

  - **Analysis environment (incl. databases) must be supported outside the LAT team**

Implications (1): Model Fitting

- **Model fitting, with models of the general form**

$$I(x,y,E,t;\tilde{\alpha}) = Diffuse(x,y,E;\tilde{\alpha}_d) + \sum_i S_i(E,t;\tilde{\alpha}_i)\delta(x-x_i,y-y_i)$$

$$\tilde{\alpha} = \{\tilde{\alpha}_d,\tilde{\alpha}_i, i=1...n\}$$

$$C(x,y,E;\tilde{\alpha},\tilde{\beta}) = \iint\left[\int_{t_1}^{t_2} I(x',y',E',t;\tilde{\alpha})dt\right]\varepsilon(x',y',E';\tilde{\beta})PSF(x-x',y-y',E';\tilde{\beta})W(E',E;\tilde{\beta})d\Omega'dE'$$

I is the model intensity distribution, which includes diffuse emission and a number of point sources, and depends on the set of parameters α.
ε is the exposure, which has implicit zenith angle cuts used to select photons, and depends on the time range and instrumental parameters β *(e.g., inclination angle, plane of conversion)*.
C is the distribution of observed photons for the time interval $[t_1, t_2]$ and depends on α *and* β.

- **The likelihood function is used to compare observed photon distributions with a model (C), usually with binning of parameters, to define confidence ranges for parameters and compare models (e.g., iteratively search for point sources)**
- **So need to retrieve γ's, calculate exposure, and look up parameter dependence of instrument response functions.**
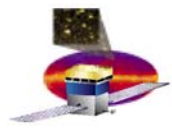
# Implications (2): Interstellar Emission Model

- **CR interactions with interstellar gas and photons make the Milky Way a bright, pervasive source**

- **Model to be developed in advance of launch**

  - **Won't need a 'good' one until real data arrive**

  - **Open issues from EGRET (e.g., GeV excess), and newly-available data relevant to gas distributions, ISRF, and CR**

  - **production and propagation mean room for improvement**

- **Will refine the model at least once after launch**

  - **Source-subtracted sky map**

  - **Limit the number of revisions for stability, reproduceability of results**

EGRET interstellar emission model (> 100 MeV)
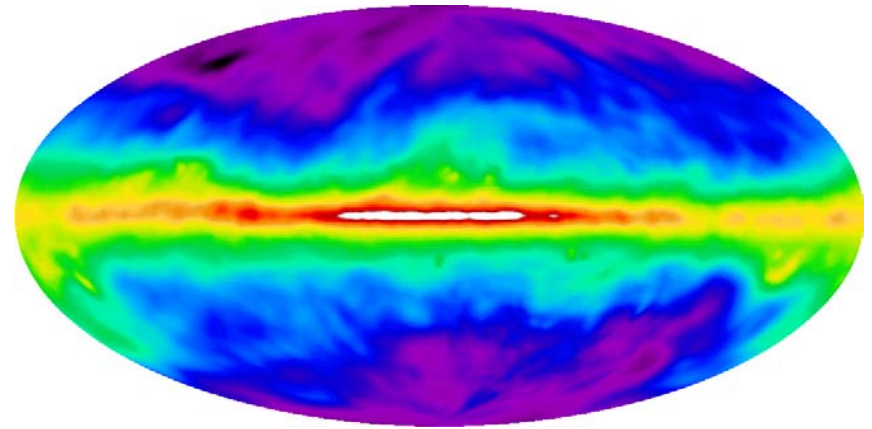
Hunter et al. (1997, ApJ, 481, 205)
Sreekumar et al. (1998, ApJ, 494, 523)
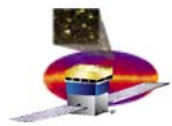
Infrastructure (1): Data Flow

- **Processing and databases**

Source Sim. is a phony Level 0 database

Low-Level (Level 1 processing)

MOC

Source Sim. Gen. — Source Sim — Low-Lev Calib — Ev Disp

Level 0 — Recon — Bkgnd Rej — Level 1

Exp Hst Extract — Timeline

Part of Data Processing Facility
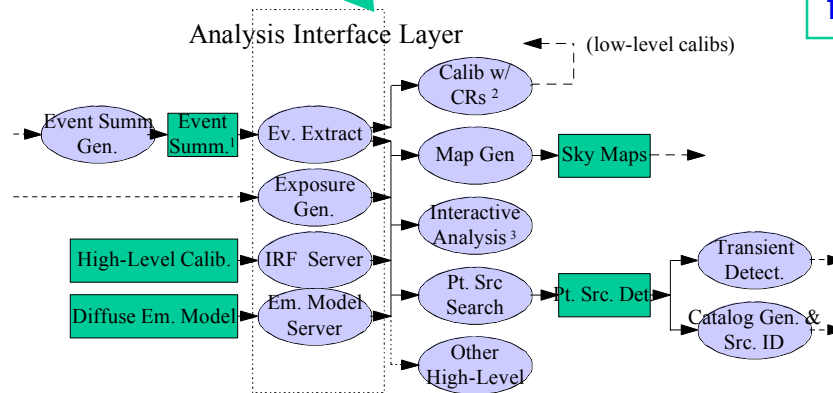
Mirror sites (within LAT team and SSC) will reproduce from here forward

High-Level

Analysis Interface Layer

(low-level calibs)

Event Summ. Gen. — Event Summ.[1] — Ev. Extract — Calib w/ CRs [2]

Exposure Gen. — Map Gen — Sky Maps

High-Level Calib. — IRF Server — Interactive Analysis [3]

Diffuse Em. Model — Em. Model Server — Pt. Src Search — Pt. Src. Det — Transient Detect.

Other High-Level — Catalog Gen. & Src. ID

[1] spatial access (celestial and instr. coords)
[2] evaluate CAL, ACD response, TKR alignment, dead strips,...
[3] spectroscopy, confused and extended sources,...

10's of Mbytes passed out of Analysis Interface Layer for a typical analysis: ~10^6 γ's + ~10^7 element exposure table

Infrastructure (2): Analysis Environment

- **Databases (Analysis interface layer and higher-level)**

- **Non-database infrastructure**
  - **Interactive (GUI) and batch (command line or script driven) interfaces are required depending on the analysis; existing candidates to be studied**
  - **Map generation**
  - **Image & plot display & printing**

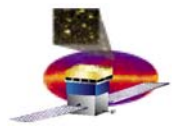| Database	Contents	Access Criteria	Used by
Event	Full info. for each event, including reconstruction (Level 1 database)	Time or event number	Event Summary constructor, event display, low-level calib monitoring
Event Summary *	Energy, direction (celestial and instr. coords), time, plane/tower/log layer of conversion, event id and bkgnd rej/quality flags	Energy, direction, time range, event flags, event ID	High-level map generation and analysis, CR event selection
High-Level Calibration *	Instrument response functions as functions of energy, angles, plane, time, …	Energy, angles, time, ...	Exposure gen, high-level analysis
Exposure History (Timeline) *	S/C position, orientation, LAT mode, and livetime for regular ~30s time intervals	Time range	Exposure gen.
Source Sim.	Monte Carlo equivalent of Level 1 data, with truth info, and run/config. ID	As Event	Recon
Point Source Catalog †	Summary of Point Source Detection, flux histories and candidate source IDs	Coordinates, spectral hardness, variability index, …	Catalog access interface
Pulsar Ephemerides	(radio) Timing parameters for known pulsars, contemp. with GLAST mission	Pulsar name	Barycenter corrector
GRB †	GRB profiles, source info.	GRB ID	Catalog access interface

* Analysis interface layer   † Higher-level ('Level 2')

Infrastructure (3): High-Level Analysis

- **Based on scientific requirements for the LAT inferred from AO and SRD**

- **Definitions of modules are motivated by discussions at LAT Software Working Group and Collaboration meetings, as well as EGRET experience within the team**

- **Exposure generation is implicit**

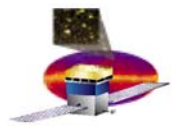| Name	Function	Inputs	Outputs
Point-source detection	Analyzing a given region of the sky for point sources	Analysis interface layer	Locations, fluxes, significances, spectrum or spectral hardness);
Point-source spectroscopy	Model fitting with flexible definition of spectral models; possibly developed as part of the general likelihood analysis capability described below (Extended sources and confused regions)	Analysis interface layer	Model coeffs and uncertainties
Extended sources and confused regions	'custom' model fitting. Interactive analysis largely will be model fitting (parametric), allowing flexible specification of source – multiple point sources, spectral models, arbitrary extended sources	Analysis interface layer	Model parameters, confidence ranges
Source variability	Detecting flares (short term, for issuing alerts), pt. source vs. extended source determination (longer term, for quantifying variability)	Point source detection database	Flux histories, estimates of variability
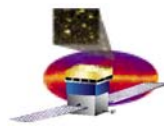
Infrastructure (4): High-Level Analysis (cont)

- **GRB & pulsar analysis is like that for point sources except for time profiles**
  - **For burst profiles and phase folding & periodicity searches, want to select events by probability of association**
- **Other tools (LOE, *not necessary for sci. req.*)**
  - **Periodicity searches**
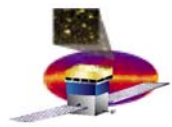 - **Multi-γ events**
  - **Polarization...**

* Also may need a tool to display times when target was in FOV to select intervals with greatest continuous coverage.

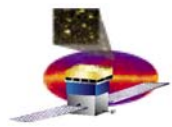| Name	Function	Inputs	Outputs
Source identification	Quantitatively defining probabilities of associations of LAT pt. srcs. with srcs. in other astronomical catalogs	Point source catalog	Point source catalog
GRB time profiles/Pulsar search event selection	Constructing time profiles for user-defined event selection criteria	Analysis interface layer (Event Summary)	Time profile histograms (perhaps normalized by IRFs, with periods outside FOV indicated), tables of events associated with a burst
Pulsar phase calculation	Assigning pulsar phases to a set of photons based on the timing params for the pulsar, to allow phase-resolved analysis for most of the analysis tasks, like spectral meas., and phase binning - for histograms and maps.	Analysis interface, Pulsar Ephemerides	Phase assignments by event number
Inflight calibration	Monitoring effective area via fluxes of pulsars, monitoring PSF via phase-selected photon distributions around pulsars.	Analysis interface	Flux histories, PSF profile plots, tables

Infrastructure (5): Observation Simulators

- **NOTE: These are not science tools per se**
  - *Low-level*
    - **For full-blown MC through Glastsim**
    - **Useful for developing and testing the DPF and SAS systems**
  - *High-level*
    - **For validating high-level analysis tools; also useful as a proposal preparation and observation planning tool throughout the mission**
  - **Depends on having an *orbit/operations simulator* with reasonable fidelity**
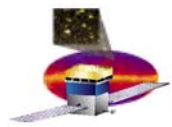
GLAST Science Support Center

- **Where the world outside the LAT team will obtain LAT data and analysis software**

- **Other roles include organizing observing/archival analysis proposal peer reviews and supporting observation scheduling**

- **Constituted early (relative to Compton GRO SSC) and already staffing up with scientists (Lead: David Band) and scientific programmers**

- **SSC will be located at GSFC, awarded in part owing to HEASARC connection (potential synergies, current ultimate archive site)**

  – **High-Energy Science Archive Research Center is NASA's archive center for high-energy (mostly X-ray) missions**

  – **HEASARC has strong ideas about data formats and analysis environments (with goals of portability and maintainability)**

Project Data Management Plan

- **From *Guidelines for Development of a Project Data Management Plan*, 1993**

  - **"In principle, the PDMP is intended to address the management of data from space science investigations, from the point of their reaching the ground, to their entry into permanent archives…."**

  - **" In practice, the PDMP will be used primarily to identify data products, and those institutional facilities used during the "project data use" phase, and [NASA] archives at the back end.... The effectiveness of the data management and archiving process depends on a well-defined relationship between the project, investigators, and the archive facility. With cooperation, NASA's data management infrastructure will continue to provide a valuable service to all and help preserve our precious space heritage through preserving the data."**
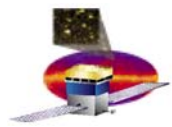
PDMP (2)

- **The PDMP defines data flow, who does what (including software and database development) in the LAT data processing stream.**

- **The PDMP is a project document, the responsibility of the Project Scientist, but must be signed off by all of the usual suspects.**

- **For GLAST, the initial version of the PDMP is essentially complete**

  - **It is receiving particular attention early on because it is also the vehicle for specifying the role of the HEASARC/SSC in the development of the higher-level analysis system.**
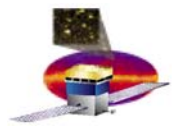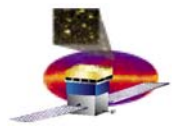
LAT Team-SSC Coordination

- **General relation with SSC is shaped by mandates from the AO**
  - **One analysis environment (LAT team + outside users)**
  - **Long-term archiving & analysis must be supported**
  - **Useability by guest investigators (stability, documentation, support)**
- **Some progress in defining the specific responsibilities**
  - **Oct 2001 – Agreement between P. Michelson, N. White, & J. Ormes (to be incorporated in PDMP and with more detail in any ICDs)**
    - **Event reconstruction software (part of the Level 1 pipeline) will not be useful or used by outside users. It does not need to conform to community standards, although it must be under configuration control, and developed with maintainability a priority, for ultimate archiving.**
    - **Level 1 and calibration data will be delivered to the SSC as FITS files. I/O from analysis tools will be via FITS format files that conform to OGIP conventions for high-energy mission data. Analysis tools for community use must be written to be portable and not dependent on proprietary software. Interfaces and software will be compatible with tools developed to support other high-energy missions.**
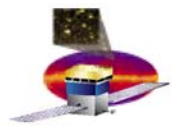
LAT-SSC (2)

- **Level 2 data products, tools, and documentation must be made available promptly, with release schedule and product specifications TBD (in PDMP).**

- **Deliverable Level 2 software will be in a configuration control system shared by the LAT and SSC. Non-core LAT team tools will not be subject to this control.**

- **LAT and SSC will jointly define the calibration files and tools for IRFs, and IRFs will be in HEASARC's CALDB format used for other high-energy missions.**

- **Some means will be provided to search the Level 1 data efficiently on consideration of the large FOV and scanning operation of the LAT. The entire Level 1 event data and associated Level 2 software shall be made available for general use. Search and retrieval of Level 1 data shall be implemented with portability and minimal dependence on proprietary software in mind.**

LAT-SSC (3)

- **SSC will participate in the software effort at all levels, including development of requirements for Level 1 and Level 2 pipelines. LAT will operate the Level 1 processing pipeline. SSC will participate in testing, but not independently develop Level 1 processing software or independently perform Level 1 processing. The SSC will have a backup Level 1 processing pipeline (to be run at NASA's discretion).**

- **Level 2 and high-level standard analysis tools (to be specified in PDMP) will be jointly developed by the IOC and SSC. SSC will concentrate on higher-level data products and support the use of multi-mission analysis tools.**

- **Interstellar emission model is the responsibility of the LAT team, and shall be delivered to the SSC.**

- **Level 0 data will be delivered directly to the SSC from the MOC and archived at the HEASARC. Level 1 processing software and calibration data will be delivered to the SSC for archiving, according to schedule to be specified in the PDMP.**
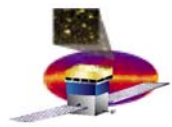
LAT-SSC (4)

– **December, 2001 agreement between LAT PI and SSC principals on process for cooperation on development of tools and databases**

- **The 'standard' analysis environment and databases first should be carefully defined (by a LAT-SSC working group) at the level of requirements documents.**

- **Requirements should also be jointly specified for, e.g., coding rules, languages, platforms, data formats, testing.**

- **Then come to some agreement re: development responsibilities. (This will in turn likely depend on the analysis expertise represented by members of the SSC.)**

- **LAT team is ultimately responsible for the high-level analysis software, with SSC part of the programming, testing pool (with software acceptance the role of the LAT-SSC working group).**

– **Other progress**

- **Data Products Working Group (June–December 2001)**

- **LAT, GBM, SSC, and Project membership**

- **Define in as much detail as practical now all of the data products to be exchanged - contents, size, delivery schedule**

- **Report to J. Ormes**
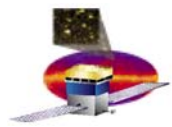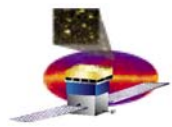
Organization for Development

- **For development of the high-level tools and databases, organization by 'subsystem' is under consideration**

- **Each subsystem would have a designated lead**

- **For all subsystems, the near-term tasks are to define requirements and interfaces**

- **Definitions of the timescales are based on overall milestones:**
  Near term ~ 1/02–9/02,
  Intermediate term ~9/02–6/03,
  Long term ~ 6/03–2/05

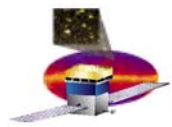| Subsystem	Near Term	Intermediate Term	Long Term
Analysis Modules	Specify requirements for the analysis modules, distinguishing 'deliverable' (to SSC) from 'internal' analysis software; study algorithms for likelihood analysis	Implement likelihood analysis; other modules	Continue implementation and testing; implementation of high-level databases (e.g., point-source catalog); Refine interstellar emission model, with 'hooks' for post-launch tuning
Analysis Interface	Event database requirements spec. & evaluation; Define database contents; Representation of high-level calibration; Exposure calculation	Post-selection implementation of database (e.g., spatial searching); detailed negotiation of contents of Level 0 data	Database mirroring; Portability of analysis interface layer
Utilities	Specify requirements for, e.g., Processing Manager and map transformations	Implement map transformations	Implement Processing Manager
User Environment	Define requirements; identify and evaluate existing candidates for interactive commanding/GUIs	Define user interfaces for the the analysis modules; Begin implementation	Continue implementation
Observation Simulation	Define requirements; interface to existing orbital dynamics model	Define model sky, and (for Fast MC) sampling according to IRFs	Refine simulation (e.g., azimuth dep. of IRFs, transient sources, pulsars)
Development Infrastructure	Evaluate adequacy, e.g., of Gaudi, coding, documentation rules; Establish CVS repository	Implement common classes for high-level analysis; Define standard testing procedure for each module	

Organization (2)

- **The work by the defined subsystems will be 'service' work for the collaboration and limited in scope to the high-level analysis tools and databases of the standard analysis environment.**

- **Other high-level tools will certainly be developed (e.g., pulsar periodicity searches, GRB pulse decomposition, high-resolution spectroscopy), but not for general use.**

Potential Backup Slides Follow -->

Interstellar Emission Model

- **A (so far) informal working group has organized to plan the interstellar emission model for LAT data analysis**
  - **Post-Gamma 2001, Baltimore**
  - **During ICRC 2001, Hamburg**
- **Practicality of 'factoring' the calculation**
  - **Defining components on a standard grid would be a starter**
  - **Also surveying the surveys (radio, mm, IR, …) available/anticipated for ISM & ISRF**
  - **Derivation of 3-dimensional distributions of ISM**
  - **Investigation of π^0 production function**
- **Model will undoubtedly be refined based on LAT observations, ideally during the first year**
  - **For reproduceability of analyses, want model stable, with infrequent revisions**