Data Flow & Leve1 1 Pipeline

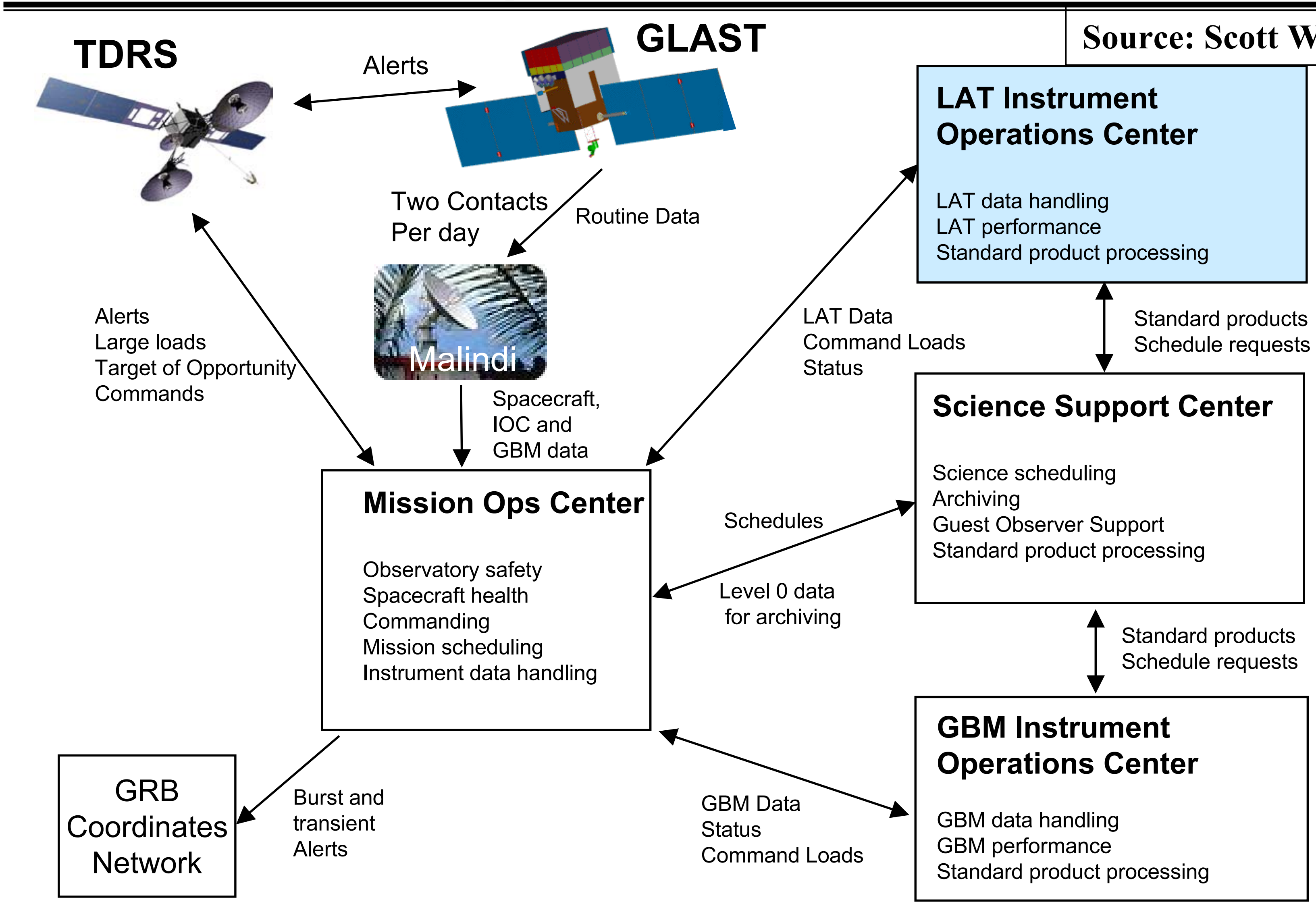
- High level specs in L3 & L4 documents:
  - Level 4 - LAT-SS-00505-01
  - Level 3 - LAT-SS-00020-01
- Pipeline Server Implementation Plan - LAT-TD-00773-01 (draft – in review)
- Database Specification - LAT-TD-00553-01

- Quick summary of Pipeline scope
  - Process real data – EM?, CU, Flight
  - Generate bulk MC
  - Facilitate calibrations (monitoring and specialty datasets)
  - Near real time diagnostics from Level 1 chain

  - Fully automated
  - Quick turnaround

Mission Operations Architecture

Source: Scott Williams



TDRS

GLAST

Alerts

Two Contacts
Per day

Routine Data

Malindi

Alerts
Large loads
Target of Opportunity
Commands

Spacecraft,
IOC and
GBM data

LAT Data
Command Loads
Status

LAT Instrument Operations Center

LAT data handling
LAT performance
Standard product processing

Standard products
Schedule requests

Science Support Center

Science scheduling
Archiving
Guest Observer Support
Standard product processing

Mission Ops Center

Observatory safety
Spacecraft health
Commanding
Mission scheduling
Instrument data handling

Schedules

Level 0 data
for archiving

Standard products
Schedule requests

GBM Instrument Operations Center

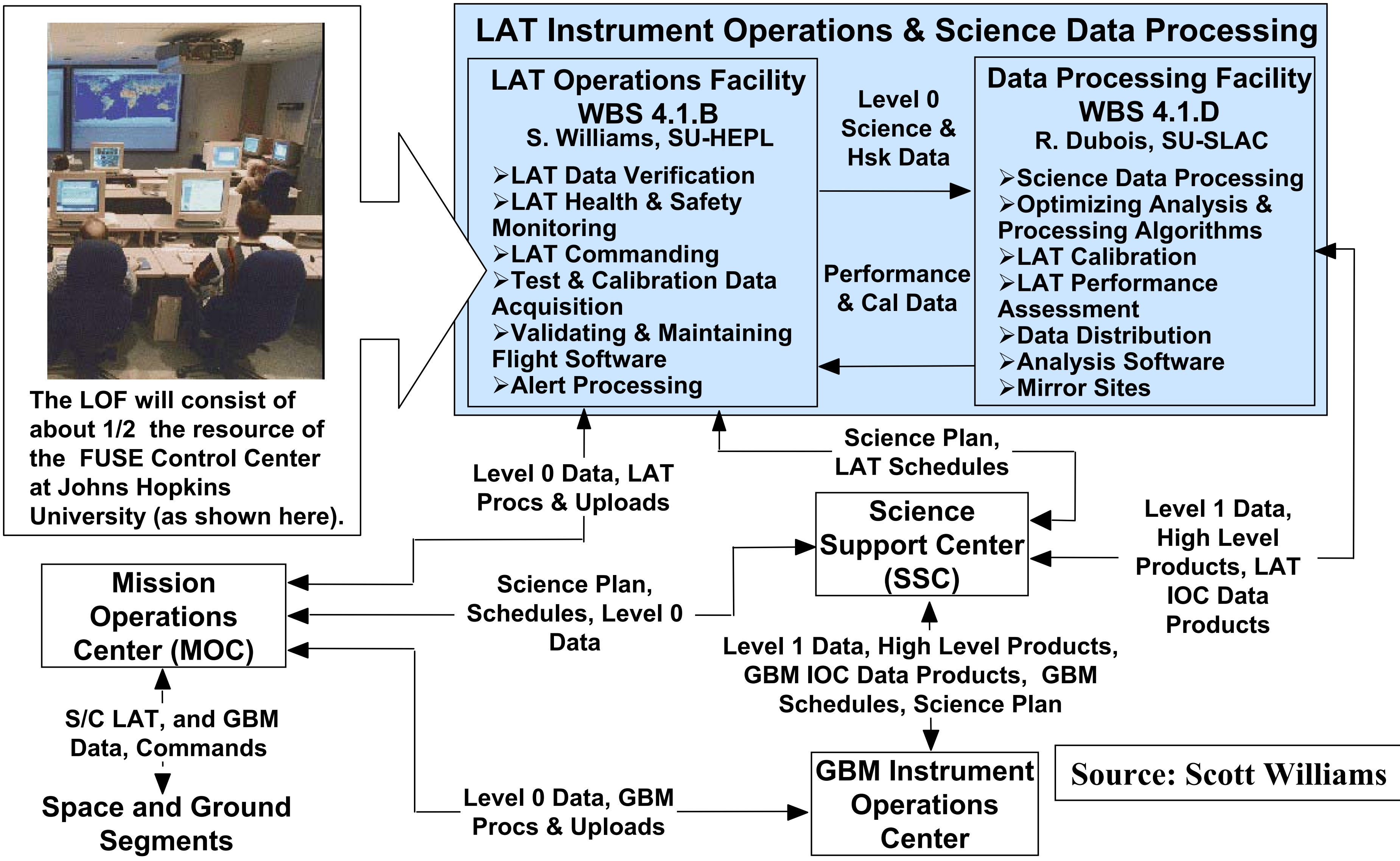
GBM data handling
GBM performance
Standard product processing

GBM Data
Status
Command Loads

GRB
Coordinates
Network

Burst and
transient
Alerts

LAT Operations Facility Functions



LAT Instrument Operations & Science Data Processing

LAT Operations Facility
WBS 4.1.B
S. Williams, SU-HEPL

- ➢LAT Data Verification
- ➢LAT Health & Safety Monitoring
- ➢LAT Commanding
- ➢Test & Calibration Data Acquisition
- ➢Validating & Maintaining Flight Software
- ➢Alert Processing

Data Processing Facility
WBS 4.1.D
R. Dubois, SU-SLAC

- ➢Science Data Processing
- ➢Optimizing Analysis & Processing Algorithms
- ➢LAT Calibration
- ➢LAT Performance Assessment
- ➢Data Distribution
- ➢Analysis Software
- ➢Mirror Sites

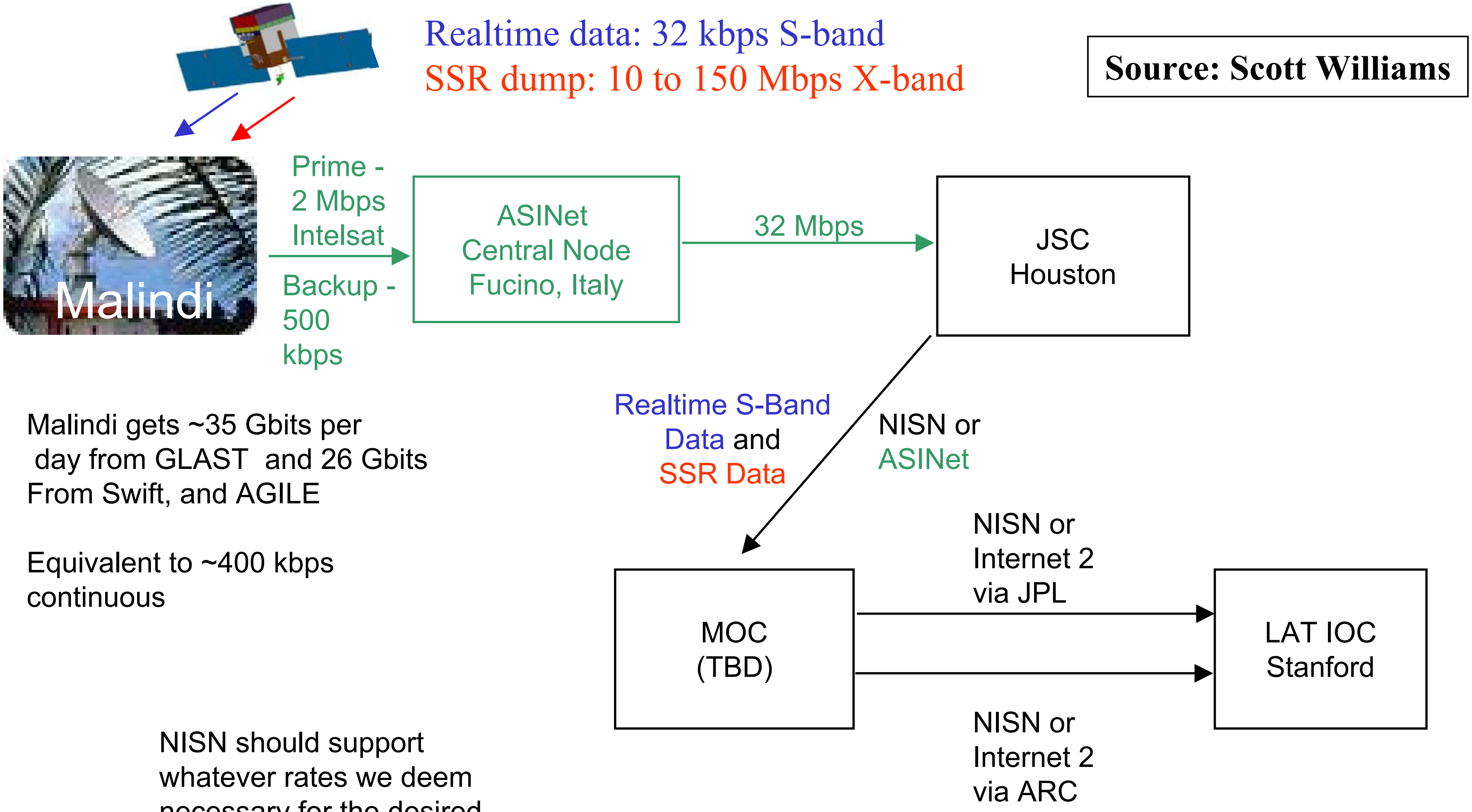
The LOF will consist of about 1/2 the resource of the FUSE Control Center at Johns Hopkins University (as shown here).

Level 0 Science & Hsk Data
Performance & Cal Data
Science Plan, LAT Schedules
Level 0 Data, LAT Procs & Uploads
Science Support Center (SSC)
Level 1 Data, High Level Products, LAT IOC Data Products
Mission Operations Center (MOC)
Science Plan, Schedules, Level 0 Data
Level 1 Data, High Level Products, GBM IOC Data Products, GBM Schedules, Science Plan
S/C LAT, and GBM Data, Commands
Space and Ground Segments
Level 0 Data, GBM Procs & Uploads
GBM Instrument Operations Center

Source: Scott Williams

LAT Data Path

Source: Scott Williams



Realtime data: 32 kbps S-band

SSR dump: 10 to 150 Mbps X-band

Malindi

Prime - 2 Mbps Intelsat

Backup - 500 kbps

ASINet Central Node Fucino, Italy

32 Mbps

JSC Houston

Realtime S-Band Data and SSR Data

NISN or ASINet

MOC (TBD)

NISN or Internet 2 via JPL

NISN or Internet 2 via ARC

LAT IOC Stanford

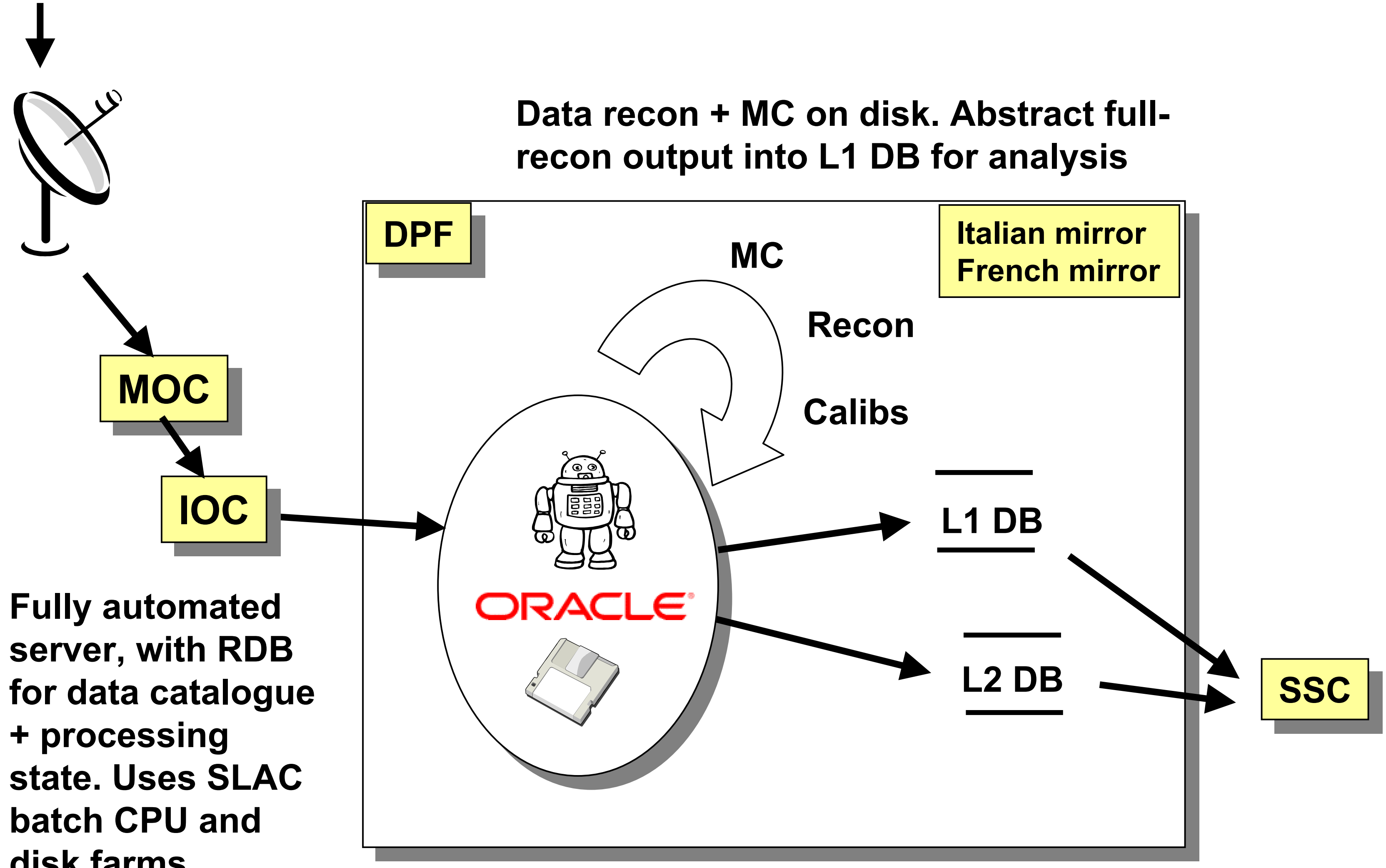
Malindi gets ~35 Gbits per day from GLAST and 26 Gbits From Swift, and AGILE

Equivalent to ~400 kbps continuous

NISN should support whatever rates we deem necessary for the desired latency Another option is Internet 2

Realtime S/C and LAT HSK Data
Followed by processed Level 0 Data

Data Flow



Data recon + MC on disk. Abstract full-recon output into L1 DB for analysis

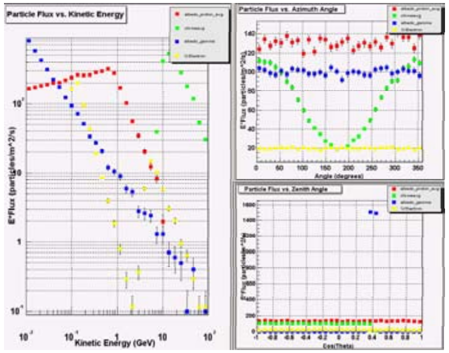
Fully automated server, with RDB for data catalogue + processing state. Uses SLAC batch CPU and disk farms.

Parts of L2 processing also automated

Instrument Simulations and Reconstruction

“Gleam”

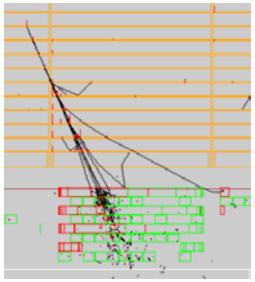
Source Fluxes



Particle Transport

Geometry

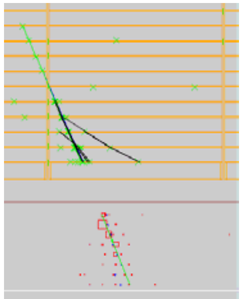
3 GeV gamma interaction



“Raw” Data

Instrument data

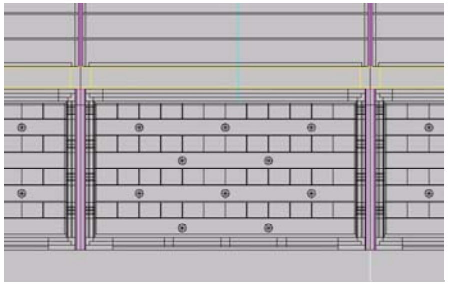
Recon



3 GeV gamma recon

Background Rejection - Particle ID

Instrumental Calibrations



CAL Detail

What is GLEAM?

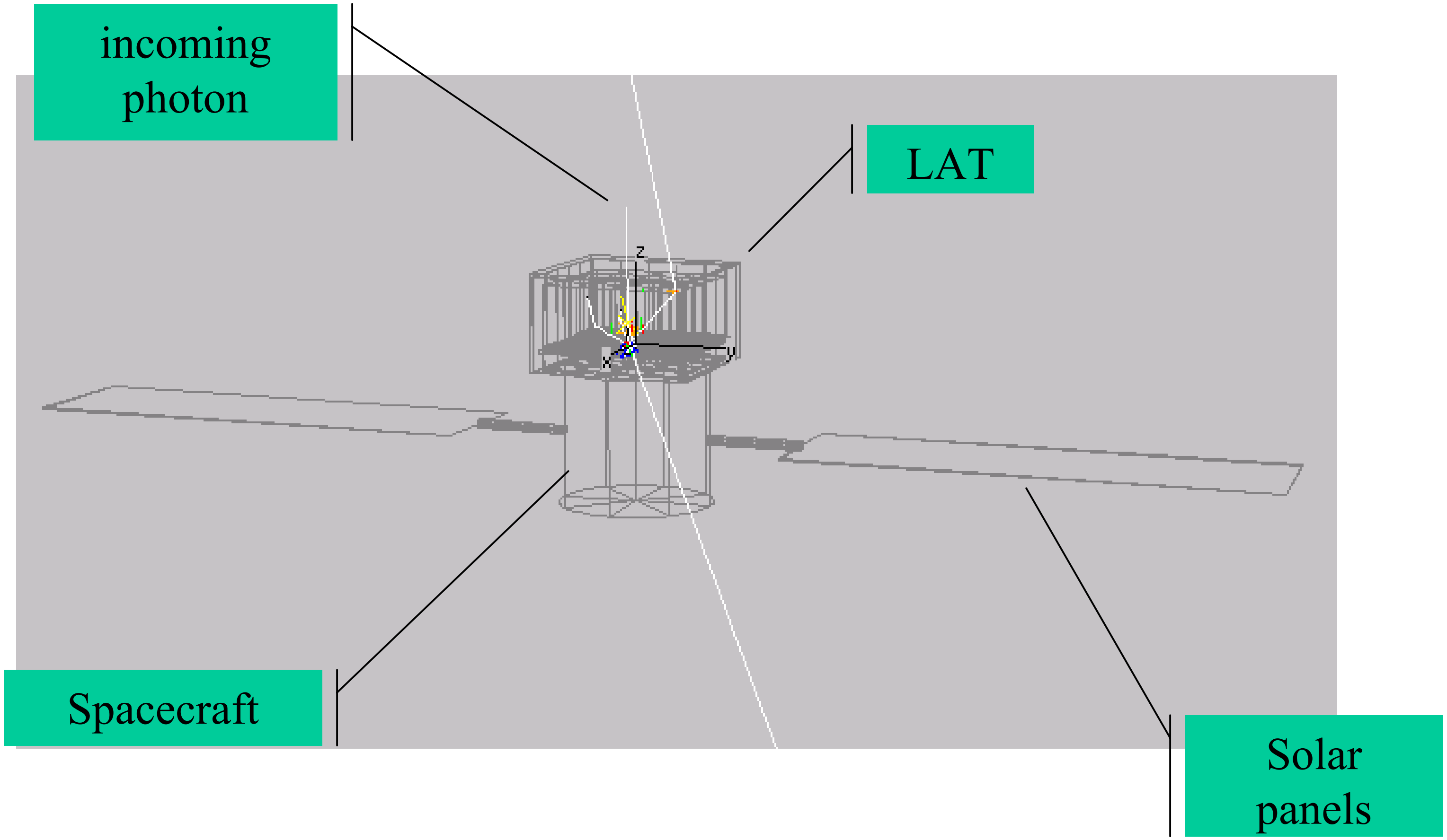
Already described by Richard, see his talk: this talk will fill in a few details.

The name (originally an acronym by L. Rochester) for an application that does everything:

- Manage an (optional) GUI interface, with a 3-D display
- Manage the detector description, including full geometry
- Manage a catalog of particle sources, ranging from simple “test-beams” to GRBs; handle geometry of the orbit.
- Run the Geant4 simulation engine to trace particle trajectories, simulating interactions, scattering, decays, etc, and record deposited energy in sensitive detectors (“hits”)
- Manage conversion of hits to “digis”, corresponding exactly to raw data collected by the detector
- Construct a “trigger” from the deposited information
- Run reconstruction algorithms to analyze strips, CsI xtal info, and ACD tiles.
- Write out the hits, digis, and recon to ROOT files for further analysis
- Read digis at least to avoid re-running simulation
- Summarize the track fit results as a PSF and Aeff

Source: Toby Burnett

Proof!



incoming photon

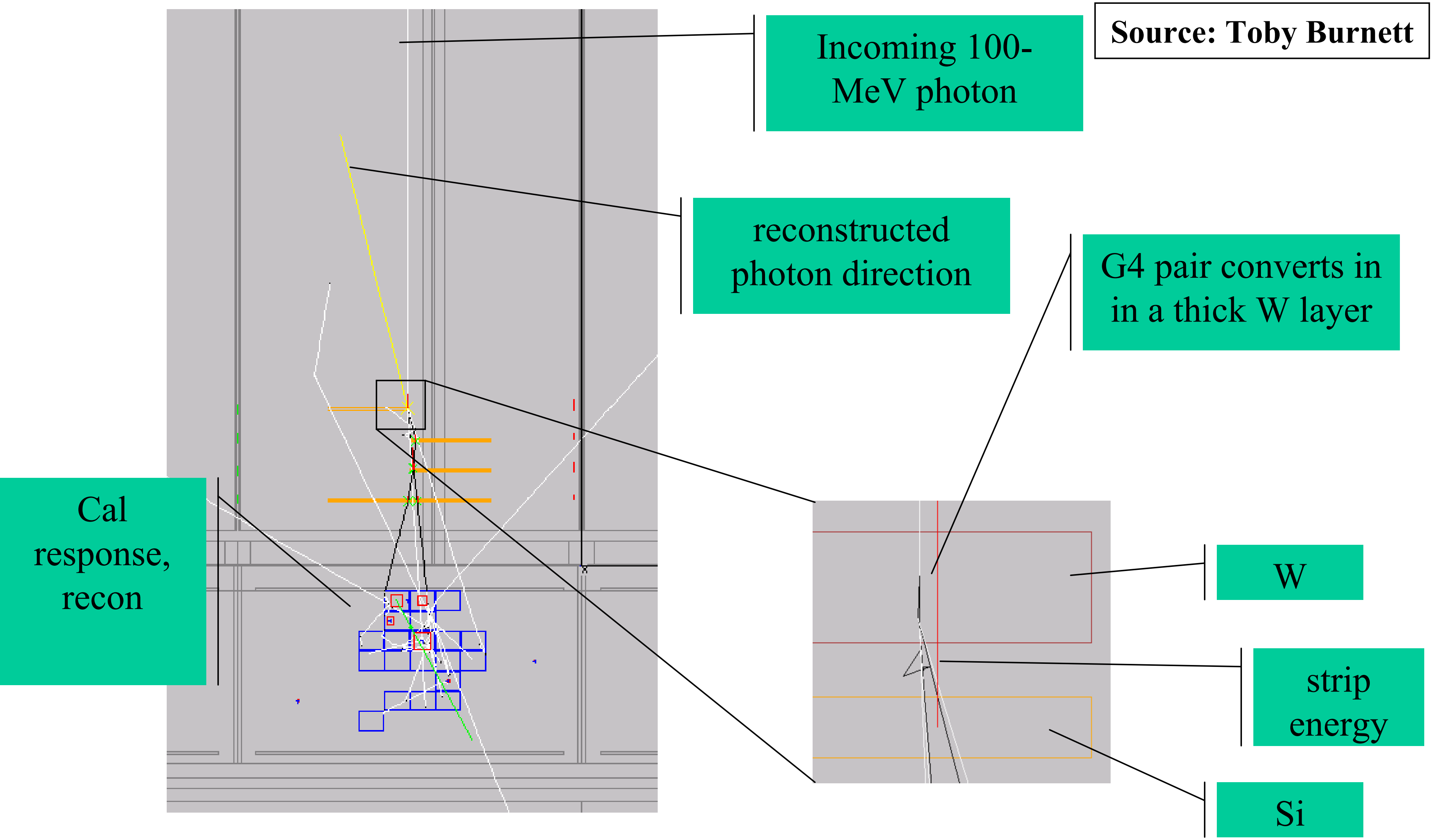
LAT

Spacecraft

Solar panels

Source: Toby Burnett

Proof, cont.



Incoming 100-MeV photon

reconstructed photon direction

G4 pair converts in in a thick W layer

Cal response, recon

W

strip energy

Si

Source: Toby Burnett

Gleam Data Products

- Current Persistent Store definition – will evolve with more experience
  - MC
    - Particle tree, Truth “Hits” in ACD, CAL, TKR
  - Digitization
    - “raw” data from TKR, CAL, ACD
    - Trigger word

Level 0.25

  - Recon
    - ACD
      - Relate tracks to hit tile(s)
    - CAL
      - Output of “clustering”
      - Position, direction from the one cluster
      - Total observed and corrected energies – leakage
    - TKR
      - Clusters of strips
      - Pattern recognition, Fit track candidates
      - vertices

Level 0.5

Upcoming Gleam Data Products

- Modifying event chain to include Event Interpretation
  - To be initiated this summer
  - Take subsystem recons as input
    - “standard” analysis – combine TKR vertices, ACD hit tiles, CAL energy/direction to synthesize event
    - Apply background and quality criteria to classify event
  - Full output will go into Level 0.5 structure

- Extract summary info for Level 1 DB
  - “summary ntuple”
  - Probably not the same as the Level 0.5 ntuple used for instrument studies

- Housekeeping to accompany science data has yet to be defined.

- Still need to deliver “header” info for each dataset

Numerology from pdrApp

- predecessor to Gleam – used for PDR performance estimate.
- will have revised numbers from Gleam this summer with iterated structures

Downlink = 3 GB/day @ 2.2M events : raw data ~ 1.5 kB/event

MC simulation takes 0.16 sec/event for all_gamma
Recon takes 0.08 sec/event

MC sim generates 1.5 kB/event
Recon generates 4 kB/event

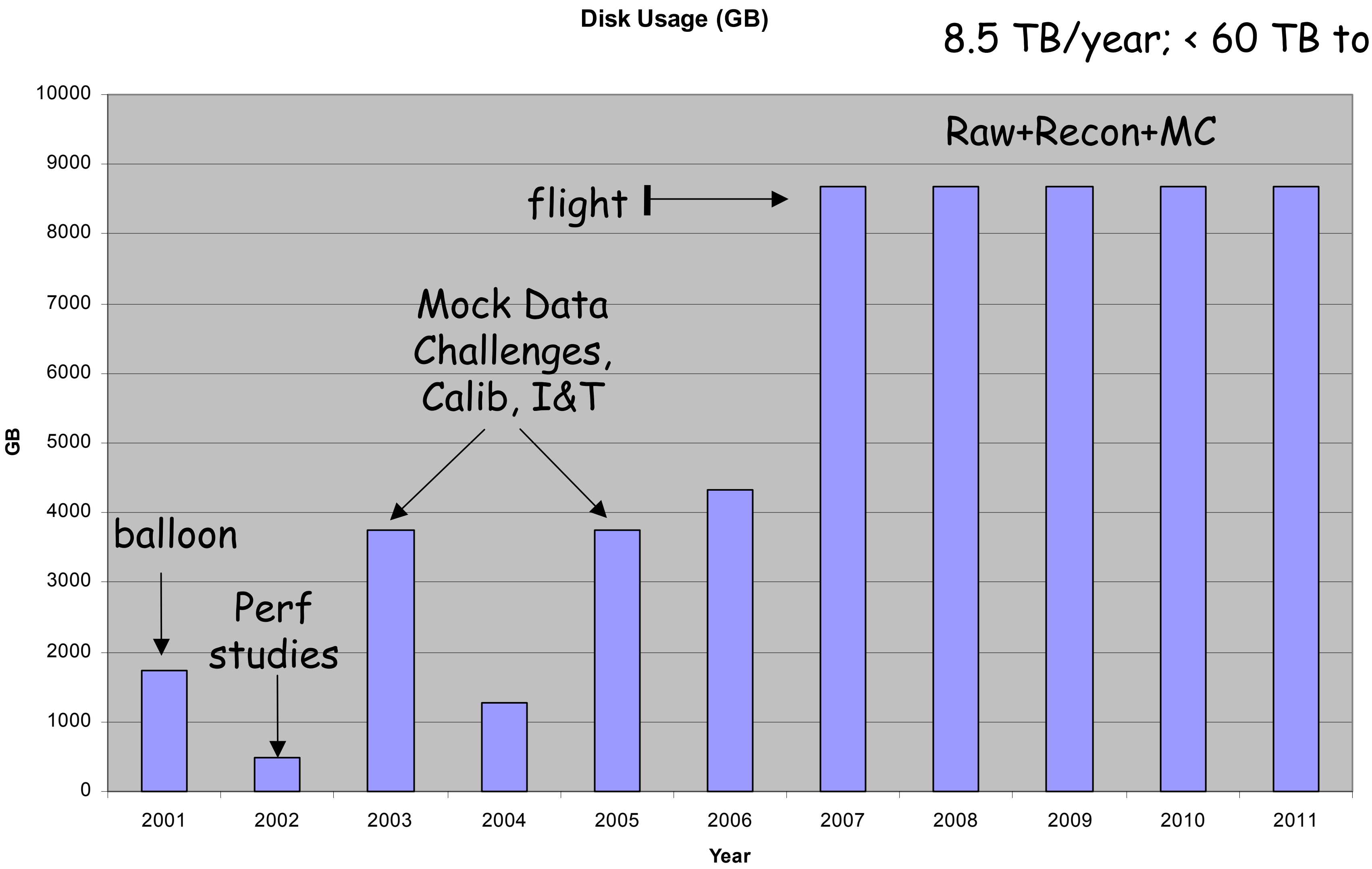
Assume equal MC generation to downlinked data rate during flight operations

A year’s re-reconstruction would add 4 TB and 100 CPU-weeks to the load

- **negotiate with SLAC Computing Center for additional resources when needed**
- **an advantage of tapping into a large batch farm!**

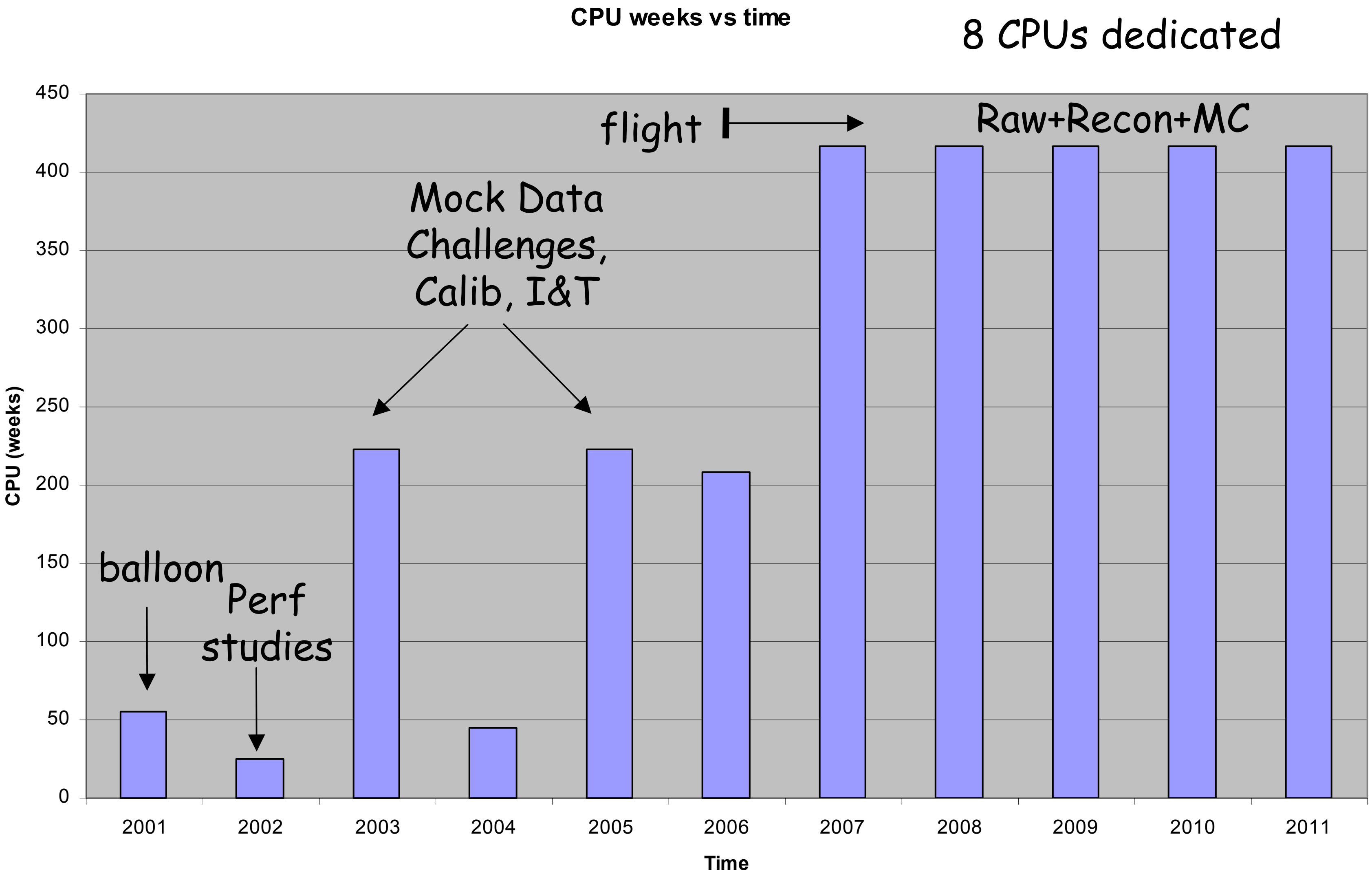
Processing Requirements - Disk

8.5 TB/year; < 60 TB total



(note: BABAR temp space is 25 TB)

Processing Requirements - CPU

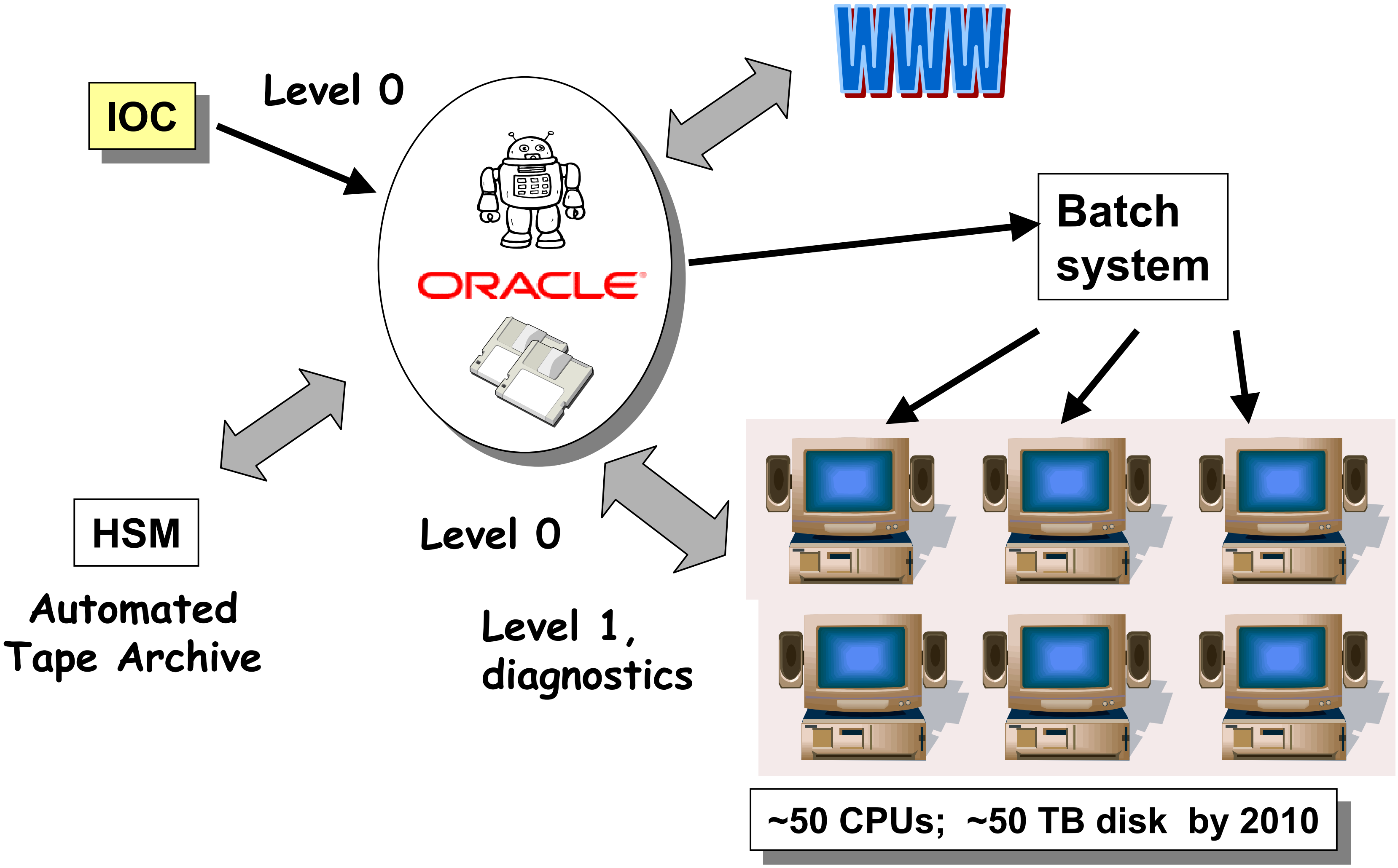


(SLAC will have ~2000 CPUs for BABAR)

Additional Pipeline Products

- Strip files (or tagged events in Level 0.5 DB) for low level calibration (peds, gains etc)
  - Select CNO etc for subsystem calibrations
  - Tag random-trigger events
  - Pipeline could run calibration algorithms if useful; humans will have to check anyhow before results put back into the system.
- Near real-time monitoring
  - Run system test diagnostics and feed back to IOC operations
  - Use full power of correlated analysis to identify problems
  - Will take form of plots and statistics, tracked in pipeline database
- Keep track of Livetime
  - Correlate livetime counters with operational modes and generate history
- Follow up onboard alerts; look for new ones

Processing Pipeline

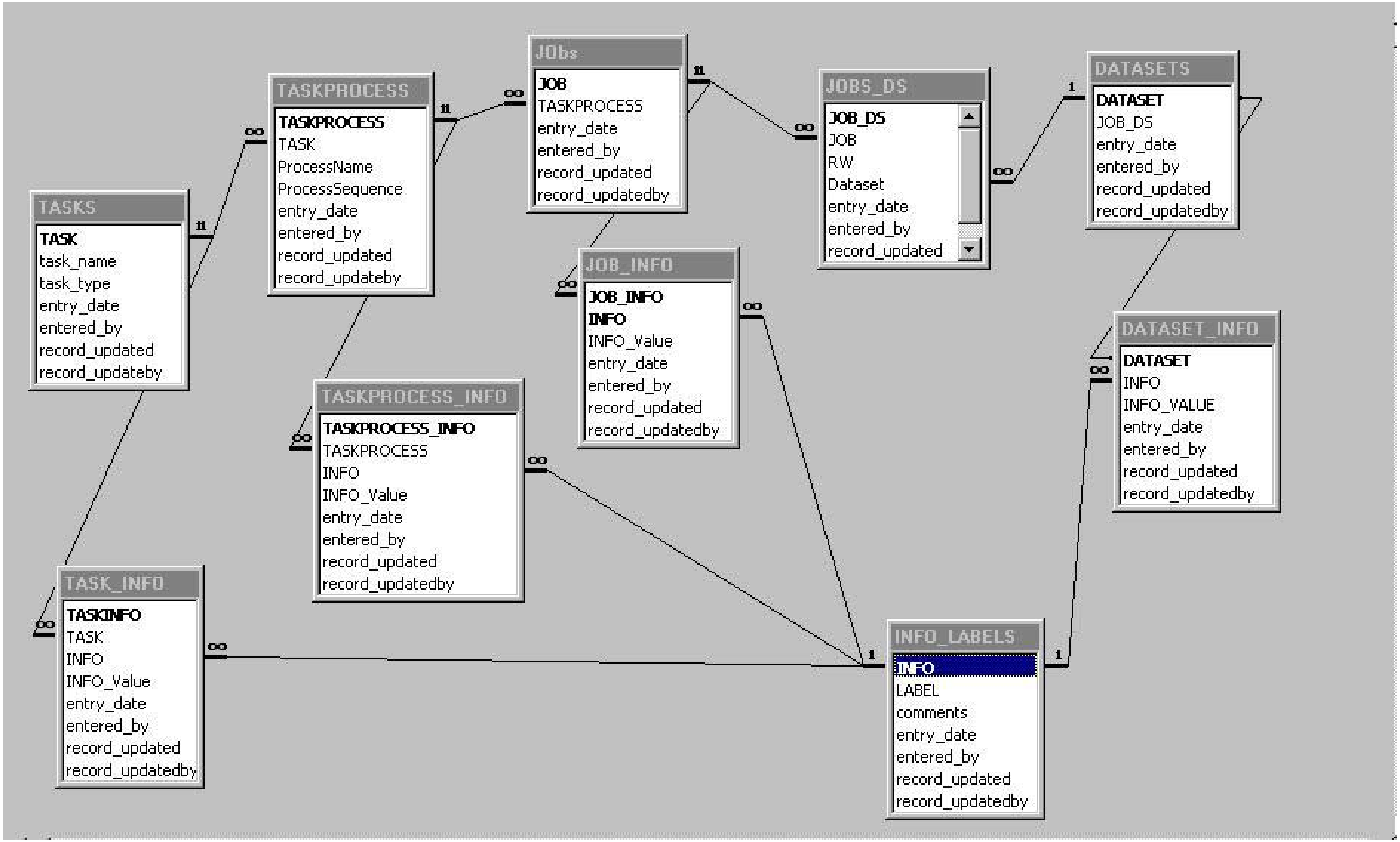


Plenty of margin on CPUs

LAT Data Processing Database

- **Heart of data processing facility is a database to handle state of processing, as well as an automated server.**
  - **relational database tracks state of file based datasets throughout lifetime in the system, from arrival at IOC or MC generation, through Level 1 output.**
  - **Automated server will poll IOC generated database entries for new Level 0 datasets and take immediate action, as well as generate MC data, and log all actions to the database.**

Prototype DB Table Layout



TASKS
TASK
task_name
task_type
entry_date
entered_by
record_updated
record_updateby

TASKPROCESS
TASKPROCESS
TASK
ProcessName
ProcessSequence
entry_date
entered_by
record_updated
record_updateby

JObs
JOB
TASKPROCESS
entry_date
entered_by
record_updated
record_updatedby

JOBS_DS
JOB_DS
JOB
RW
Dataset
entry_date
entered_by
record_updated

DATASETS
DATASET
JOB_DS
entry_date
entered_by
record_updated
record_updatedby

JOB_INFO
JOB_INFO
INFO
INFO_Value
entry_date
entered_by
record_updated
record_updatedby

DATASET_INFO
DATASET
INFO
INFO_VALUE
entry_date
entered_by
record_updated
record_updatedby

TASKPROCESS_INFO
TASKPROCESS_INFO
TASKPROCESS
INFO
INFO_Value
entry_date
entered_by
record_updated
record_updatedby

TASK_INFO
TASKINFO
TASK
INFO
INFO_Value
entry_date
entered_by
record_updated
record_updatedby

INFO_LABELS
INFO
LABEL
comments
entry_date
entered_by
record_updated
record_updatedby

LAT-TD-00553-01 – if you’re interested read this

Data Manager Prototype

Existing Data Manager Prototype is set of perl scripts that:

- **Performs automated MC batch processing using LSF and SLAC batch farm (e.g. produced 50 M background events, 10 M gammas for PDR studies)**
- **Provides utilities for processing, filtering , and displaying results of MC runs**
- **Provides very preliminary scripts for entering results of MC runs into Oracle tables**
- **Will evolve into Data Manager for Data Processing Facility (DPF) by being split into a server and set of utility packages as described in Data Manager spec**

Key Pipeline Concepts

- Media
  - Everything on disk
- Input to the Pipeline
  - FITS files appear on disk; tagged in DB
- The Processing Chain
  - Elements configured in the DB
  - Customize to task (real data, MC, etc)

LAT-TD-00773-01 (in prep)

- Communication with Processes
  - Interface hiding batch system details so that a non-SLAC implementation can use a different mechanism
- Communication with DB
  - Server does all communicating with DB, which is standard SQL
- Communication with Server
  - CLI, gui
  - Authenticate users
- Archive all files to tape

Additional Details

- Current plan is to export Level 1 db to SSC
  - Several ways to “mirror”
  - Haven’t started talking to LAT mirrors yet – still early in the game
  - Pipeline database knows about all datasets, so any mirroring should be easy
- DB keeps track of all details of code used to do processing + parameters used
- Can also be used for Level 2 automated processing
- Planning & Implementation
  - DB spec and Implementation plan to be reviewed by LAT-SSC working group
  - Alex Schlessinger targeted to implement pipeline
  - Starting to think of implementation details already
  - Try to have something ready for EM and for CDR MC generation
- On the TODO list: handle burst alerts & tracking livetime