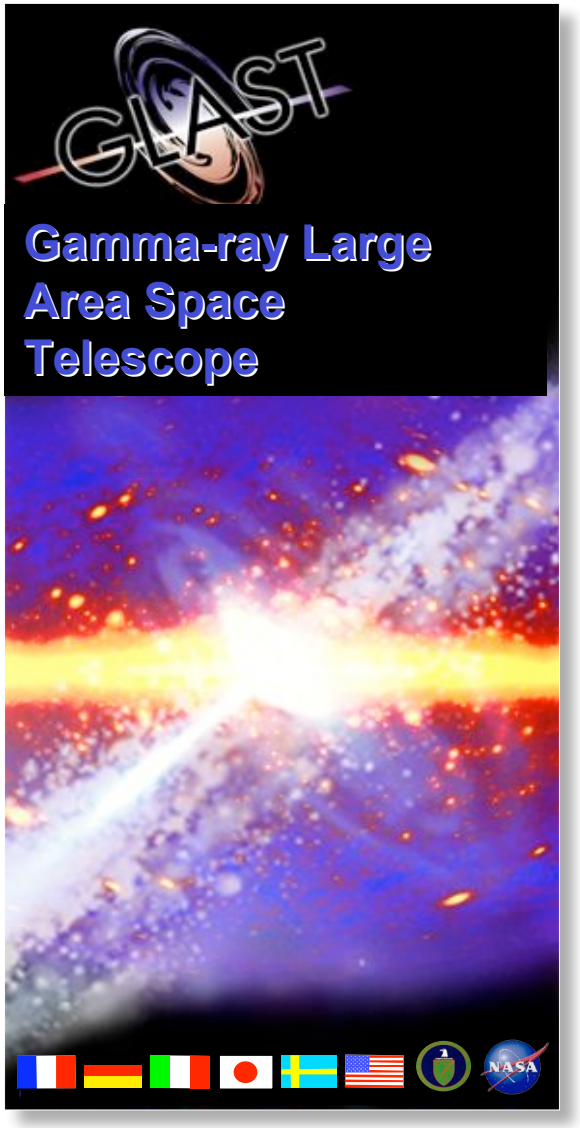


GLAST Large Area Telescope:

Service Challenge Update and Collaboration Compute Needs

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

Outline

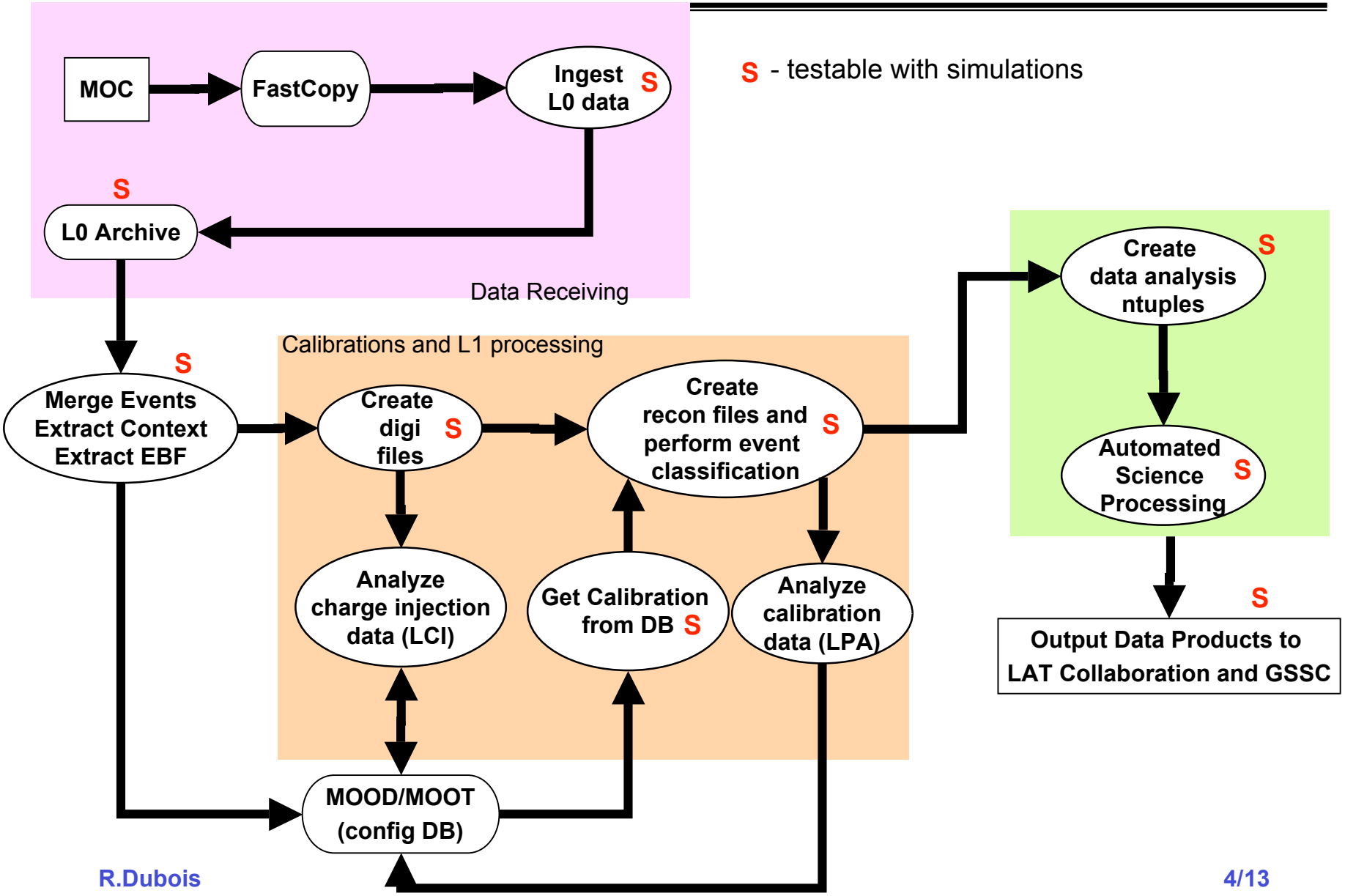
- Service Challenge update & Major Milestones
- Computing Resource Projections
- Manpower Needs

Connection to Science Groups

- Several datasets have been identified for the Science Groups use:
  - 1 yr obssim survey run (one already done last December)
  - 55 day Gleam run – ibid
    - Earth occultation not currently in exposure calculations
    - ARR
    - New orbit
    - LAT/SC misalignment
    - Background interleave for pointed observations
  - 1 year Gleam run
    - Potentially huge backgrounds run needed
    - Targeting use of Lyon, Italian farms
  - Plus a few smaller scale specialty runs as needed

- Possible timescales are mid February, end February and June for these 3 datasets

Simplified Diagram for ISOC Data Flow



Connection to ISOC

- **Rethought strategy after organization of Science Ops and Service Challenge workshop**
  - **Use ETEs for control room type functions**
    - **Shift log (eLog), L1 pipeline, Data Catalogue, Monitoring**
  - **Use simulations to prep for ETE & provide realistic science data, extended running**

- **Simulate L0 science data**
  - **Prep for ETE1 L1 pipeline usage (air shower gammas)**
  - **Realistic science data for response distributions, resource usage, latencies etc**
  - **Downlink simulations for instrument readiness tests, such as calibrations, failed sensors etc.**

- **55 day orbit run**
  - **Extended run to test ASP**
  - **Time trending of instrument quantities**

Rough Timeline

ETE1 (end Feb)

- L0, L0.5, L1 pipeline definitions
- Shift log(*)
- Monitoring - digi, recon reports à la I&T

55 Day Run (end Feb)

- Revised background interleave
- Revised spacecraft geometry
- Data Catalogue (*)
- Astro & Instrument Data Servers
- Xrootd data access (*; in use by BABAR)
- Pipeline 2 extended to Lyon

Big One Year Run

- P2 extended to Bologna
- Data Portal

Plus:

- Partial and full downlink runs post ETE1/55 day runs
- Reprocessing test

(*) - new tools

Current Compute Resources

Starting 3rd year of projected annual $300k Capital Equipment Projects

- Supplying, batch farm disk & CPU, as well as dedicated servers
- Optimize purchases based on best deals SCS can come up with

- **100 TB disk (14 remaining!) –**
  - LAT Commissioning (~60)
  - DC2/SC (10)
  - Beamtest (5)
  - Pre-ship review (5)
  - Infrastructure needs (code builds; system tests; user disk)

- **Tremendous use of SLAC Batch farm**
  - 160 cores (40 dual core dual CPU boxes)
  - Aiming for 400 cores at SLAC
    - Not looking good for quad core CPUs to be available this year

Known Liens on Resources

- Data taking at Spectrum Astro
  - Pre Christmas test rate has tailed off.
  - Expect large hit on disk in spring with 6 weeks of TVAC
    - ~24 TB (more than we have left)
- Beam Test
  - Appears to be done
- SC L0 run in February
  - 3 days running
  - ~300 GB disk
- ETE1 12 hrs data
  - small

Resource Rules of Thumb

- 1 orbit-day bkg takes 4 days to create and 2.5 TB of disk @ ~300 CPUs continuous
- 55 day sky simulation takes the same amount of time as the 1 day background sample!
- 1 yr run with 5 days' background would be

5x(4 days, 2.5 TB) + 6x(4 days, 2.5 TB) = 44 days, 27.5 TB

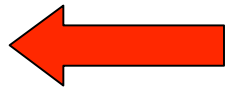
Planned Acquisitions

- SLAC
  - Order for 50 TB disk in process (paid for by IFC FY06 $ is plan)
    - Latest ship date is end Feb (from Sun)
  - $300k Capital money from SLAC to divide between disk and CPUs
    - Original plan was to wait for quad core boxes and buy 30 to get us to 400 cores before launch; rest goes to disk (~50 TB)
  - IFC has 'pledged' 80 TB/$300k for FY07
- Lyon
  - Berrie has requested 100 CPUs, 50 TB disk
    - "no problem"
    - Need manpower to help port Pipeline2 to Lyon (ClaudiaL at Montpellier has just started)
- CNAF
  - Francesco has submitted proposal for 100 CPUs, 25 TB (in 07)
    - Approved
  - May be required to access via GRID tools
  - Have not identified Claudia-equivalent yet

Future Storage Options: Orbit Data

- 10% Solution:
  - 60 TB/yr @ ~$240k for disk; ~$50k tape (or $120k if all recon kept on tape)
  - Assumes we can safely cut 90% of the background - we don't have the cuts determined yet

- Full disk in '07; 10% in '08 and onwards (add 25 TB contingency to 150)
  - ~175 TB overall @ $700k for disk; $140k for tape
  - We'll see how disk usage goes with some experience under our belts in '08



Best Plan is to optimize (175 TB, 400 cores) with ~$600k IFC+CEP as prices dictate

- No recon on disk
  - ~30-40 TB/yr @ 120-160k$ disk; 25-35 k$ tape
  - suffer latency in retrieval when needed
    - 1 TB retrieved in 1 day currently
    - Affects reprocessing merit; event displays; calibrations
    - Is there some need we'll find for fast turnaround that would make us regret not having things right there? Don't know yet.
- Wrinkles
  - Price of disk will continue to drop. How soon and how much? 20% last year
  - SCS has told us not to worry about tape costs (yet)

SAS Manpower Needs

- Already working closely with the collaboration – 20 of our ~30 FTEs come from outside SLAC!
- Could not be done without this group-wide effort
- Quite a bit of ebb and flow in who does what

- Historical contributions outside SAS/SLAC:
  - Italy:
    - Core tools, G4 support – 2,5 FTE
    - TKR alg development – 1 FTE
    - SciTools development – 2,5 FTE
    - BT release management - 0,5 FTE
  - France:
    - Core tools: 0,5 FTE
    - CAL sim/recon/calibs – ~5 FTE
    - Source Catalogue – 3 FTE
  - US:
    - Core tools – 2 FTE (UW, Goddard)
    - CAL sim/calibs – 2 FTE (NRL)
    - ACD – 1 FTE (Goddard, SVAC)
    - TKR – 1 FTE (UCSC)
    - Sci Tools - 3 FTE (GSFC)
    - Sundry contributions

- Responsibilities outside SAS/SLAC:
  - Instrument code release manager - UW
  - SciTools release manager - GSSC
  - Builds manager – Goddard
  - System tests manager – Goddard
  - Documentation - Goddard
  - Code build tool (CMT) – LLR/France
  - Build tool gui & Event display – Udine/Italy
  - Source Catalogue – CEA/France
- Responsibilities inside SAS/SLAC:
  - Overall management
  - Data handling
  - Calibrations infrastructure
  - Overall Reconstruction; TKR
- Shortfall:
  - ~ 4 FTE for infrastructure work
  - Hardest to get assistance here

What to Take Away

- **ServiceChallenge + ETEs**
  - Being used to hone the tools, complete development and test end-to-end operations
- **Prudent approach is to:**
  - Acquire ~400 cores available to GLAST
    - Lesson learned from 5-ring circus of DC2, BT, EM
  - Keep full event details on disk in '08 : ~175 TB
    - Hard to do both these without IFC contribution of 50 TB in CY06, 80 TB in CY07
    - Would shortchange CPU if push comes to shove
  - GLAST will do better science the more compute power it has access to
    - Have not hit the plateau yet!
    - Extend Pipeline 2 to France and Italy
- **SAS manpower is barebones in core software**
  - Expect to get 1 new FTE at GSFC via Steve
  - Short 3 with no prospects of getting them from the rest of the collaboration