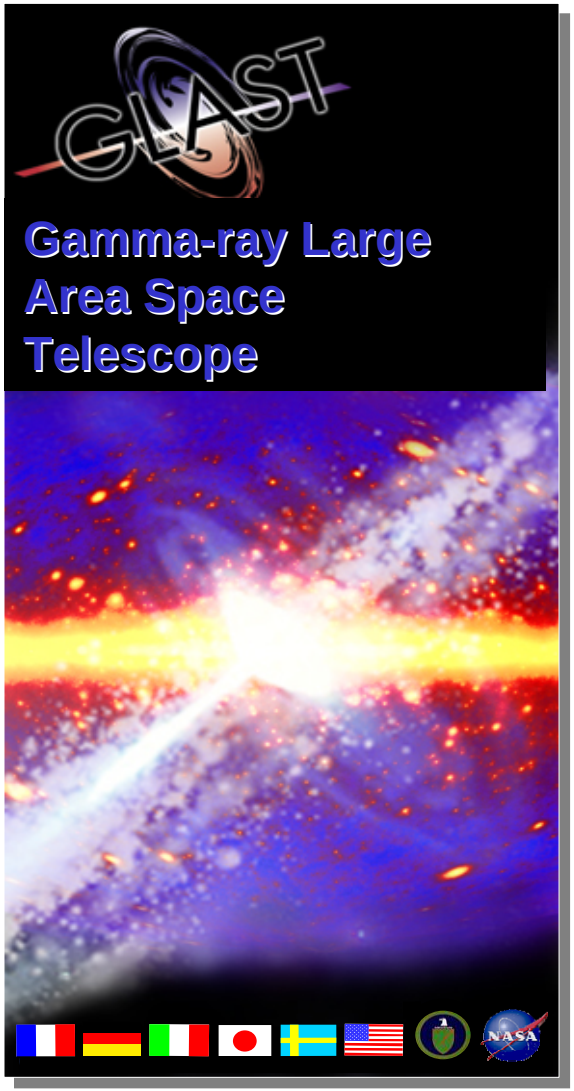


GLAST Large Area Telescope

LAT Reboot Resolution Team

February 08, 2007

Monthly Status

Erik Andrews
Jana Thayer
Pat Hascall, Joe Cullinan

RRT Status

- **Reboot summary data are maintained on the ISOC / FSW Website:**
  - **http://confluence.slac.stanford.edu/display/ISOC/FSW**
- **Update on fishbone analysis – Next pages**
- **B0-8-0 + B0-8-1 Status**
- **Diagnosing reboots during observatory test**
  - **Memory dump procedure defined**
    - **Pre-defined memory locations are dumped (~2 hours)**
    - **Turning procedure into a "blue sheet"**
  - **FSW on call 24/7 to diagnose reboots**
    - **Additional dumps may be defined after some analysis to be gathered in advance of FRB (subject to Obs I&T impacts)**
  - **FRB On-call team identified. Process produced good results as used.**
    - **Phone #s distributed and available for operators**

Fishbone Status

- The fishbone analysis has been an effective tool used to focus the attention of the team on the most likely causes
- Two major branches have been designated as not credible causes
  - Held two telecons with Fred Huegel to discuss potential hardware causes and how they could be eliminated
    - Included hardware component failures and environmentally induced failures
  - The result of these meetings was that the best description for these two classes of failures was “not credible”
    - As examples, these categories would require parallel failures in 5 boxes, would require marginal performance that was not affected by the environmental test sequence, or would require some other condition that was considered not credible
- The abbreviated fishbone charts attached present the possible and unlikely items, with non credible items removed.

Abbreviated Fishbone

Cause	Discussion/Rationale	Status
1. Hardware failure induces reboot	Identical pre-existing, intermittent or environmentally induced failure of isolated part or board in 5 independent processors is not a credible cause for reboots	Not Credible
2. Software induced reboot		Possible
2.1. Operating system flaw		
2.1.1. Priority inversion	Code has been designed to avoid this issue	Unlikely
2.1.2.OS does not provide memory protection	Code analysis performed to eliminate potential memory overwrite errors, but see 2.2.1	Very Unlikely
2.2. Application software bug		Possible
2.2.1. Memory overwrite		Very Unlikely.
2.2.1.1. Generic overwrite	Potential for overwrite documented in FSW-823, FSW-831. Static code analysis tools utilized by IVV to check for unprotected memory writes. Issues found in JIRAs xxx, xxx and resolved.	Very Unlikely.
2.2.2. Interrupt locks	Code has been designed to avoid this issue	Unlikely
2.2.3. Task exception	Could not cause watchdog timeouts, since the CPU takes an exception first. Task exceptions are handled in the exception vector. System designed to collect information on the task that attempted to execute illegally. This data is captured preserved. These resets, when they occur are captured and provide good insight into what's happened.	Possible
2.2.4.Race Conditions	Where race conditions are a potential, system designed to sequence events thru use of semaphores	Unlikely

Abbreviated Fishbone (Continued)

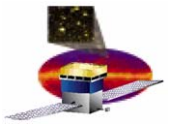
Cause	Discussion/Rationale	Status
3. Operations/environment	Not a credible cause – environmental testing was successfully completed with no change to reboot rate during any environment	Not Credible
4. LAT software interacts with computer firmware/operating system feature		Possible
4.1. Feature documented in vendor errata sheets	General note: examined errata from vendor, including newly disclosed features.	Unlikely.
4.1.1. Errata 15	LAT is susceptible, documented in FSW-820, FSW-821. Recommended work-around implemented in Build 0.8.x, so now considered unlikely. Will not rule out yet.	Unlikely.
4.1.2. Errata 24	LAT is susceptible, documented in FSW-822, FSW-824. Recommended work-around implemented in Build 0.8.x, so now considered unlikely. Will not rule out yet.	Unlikely.
4.2. Undocumented and previously unknown errata		Possible
5. EPU/SIU hardware design flaw		Possible
5.1. LCB FPGA error		Possible
5.1.1. LCB incorrectly writes memory	Writes to random areas in memory could cause an exception reboot, but very unlikely to cause a watchdog timer reboot since corrupted memory would be more likely to cause computation errors or exceptions rather than causing a process to hang.	Possible

0.8.x

- Applied workarounds for BAE errata (B0-8-0)
- Implement FSW changes resulting from code review
- In case of watchdog, use expanded LSW trace, coherent copy of interrupt stack, task stack, and dump of last ISR and subroutines called from ISR to determine whether problem is an interrupt, code corruption in an interrupt routine, or code corruption elsewhere
- In case of an exception, task ID, PC, etc are recorded
- Data to FES conversion tool being implemented in order to run LAT data taken prior to reboot through the testbed to try to reproduce the reboot

Reboot Summary (since 0.8.x)

- **Week Of 1/22: 0.8.0 loaded to lower bank of SIU0, EPU0, EPU1.**
  - **Only loaded to single bank of three processors due to upload time duration. Knowing this was a transient build.**
  - **1/26 – 1/27: Two shifts of muon runs with B0-8-0 yielded three processor resets**
- **2/1: Install B0-8-1 to catch the decrementer in a lost interrupt situation.**
- **2/6 – 2/7: Very Successful. Multiple (3) processor resets.**
- **2/8 (Today) – Found MCP750 documentation substantiating the current hypothesis.**
  - **Planning to meet with BAE to discuss later today.**
- **If analysis proves correct, the workaround is straightforward and can be part of B0-9-0. More build plan information in the FSW charts..**



Restrictions on Use of board resources

- **From the MPC750 User Manual**
- **C.13 Performance Monitor (Chapter 11)**
  - The performance monitor of the MPC755 functions the same as that of the MPC750, and is completely described in Chapter 11, “Performance Monitor,” except that for both the MPC750 and MPC755, no combination of the thermal assist unit, the decrementer register, and the performance monitor can be used at any one time. If exceptions for any two of these functional blocks are enabled together, multiple exceptions caused by any of these three blocks cause unpredictable results.

RRT Plans

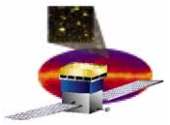
- **Plan forward:**
  - **Confirm exception conflict between the decrementer register and the TAU exceptions exists on the BAE board.**
    - **If so, Propagate finding to spacecraft (and other GSFC programs)**
  - **Re-focus on watchdog timeouts, should/when they continue to exist. Data captured by LSW will greatly assist in tracking down further issues.**
  - **Implementation of the VxWorks in a write-back / write-thru commanded option to support diagnostics/troubleshooting requirements in more real-time. (JIRA 857).**
  - **Document process & results.**

GLAST Large Area Telescope

Monthly Mission Review

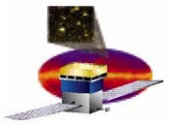
Backup

Stanford Linear Accelerator Center



“Lost Decrementer Interrupt” (1/2)

- **Immediate cause is VxWorks “kernel work queue overflow (panic)” – too much work to be completed at next process-scheduling event.**
- **Suspected root cause: erratum in RAD-750 processor chip causes failure to request an interrupt when decrementer overflows – hardware waits another 2**32 ticks of the 8.25 MHz clock (520.602 seconds) and generates interrupt on the next decrementer overflow.**
  - **Related to previous reboots**
  - **Some predictions based on this hypothesis confirmed by data from previous reboots**



"Lost Decrementer Interrupt" (2/2)

- **Plan forward**
  - **Confirmation of root cause required**
    - **Add information to trace to identify whether decrementer overflows**
    - **Install patch (B0-8-1) to identify root cause on LAT, available 2/1**
    - **Identify smoking gun**
  - **If confirmed, then implement work around (timescale 2 weeks)**
    - **Suggested work-around: replace decrementer by spare programmable timer in bridge chip to generate timing interrupts.**
      - **Uses asynchronous external interrupt path rather than decrementer interrupt.**
      - **Lose two bits of timing resolution – LSB goes from ~120 to ~480 ns.**
      - **Possible additional jitter of ~120 ns (full width).**

Process: Reboots during Observatory Test

- If a reboot occurs during Observatory Test:
  - Planned LAT testing is terminated
  - Memory dump procedure is run by LAT operator
    - Pre-defined memory locations are dumped and received
      - SIU dumps: via primary boot 1553 housekeeping*
      - EPU dumps: via diagnostic telemetry
    - Approx 3 Mbytes and two hours for this dump
    - Turning procedure into a “blue sheet” and incorporating changes due to inaccessibility of SSKI rack
  - FSW may define additional dumps after some analysis
  - FRB will be convened
  - FSW on call 24/7 to diagnose reboots

*If dropouts in 1553 housekeeping persist, our ability to diagnose SIU reboots will be impaired. If the problem cannot be fixed, can we request a dump of the housekeeping partition of the SSR to get the data dumps?

- Neil has circulated a document describing the process and timescale
  - LAT dump activity can be delayed until convenient in observatory test sequence – only impact is continuous primary boot telemetry.
  - LAT power configuration cannot be changed until the dumps are complete else data will be lost.
  - LAT reboot diagnosing activities may extend beyond planned shift boundaries and potentially impact Observatory test activities, else diagnostics will be lost.