

P116 Reprocessing Notes

P116 Reprocessing

status: **Complete**

last update: 30 November 2010

This page is a record of the configuration for the P116 reprocessing project, rebuilding FT1 files (and friends) for all of science data with new, extended event classifications.

- [P116-FT1](#) - this [task](#) will read existing (Level1) MERIT files and produce FT1 (photons) + LS1 ("fat" FT1) files

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v116.

XROOT location and file naming

Location template:

```
/glast/Data/Flight/Reprocess/<reprocessName>/<dataType>
```

Locations for P116:

```
/glast/Data/Flight/Reprocess/P116/ft1
/glast/Data/Flight/Reprocess/P116/electronft1
/glast/Data/Flight/Reprocess/P116/ls1
```

File naming:

Data Type	aka	Send to FSSC	Naming template
FT1	LS-002	Yes	gll_ph_p<procVer>_r<run#>_<version>.fit
LS1	LS-001	Yes	gll_ev_p<procVer>_r<run#>_<version>.fit

Note: 'procVer' is a field added to the file name (and the keyword "PROC_VER" in the primary header) added to the FFD 5/12/2010. Ref: http://fermi.gsfc.nasa.gov/ssc/dev/current_documents/Science_DP_ICD_RevA.pdf

Example:

```
/glast/Data/Flight/Reprocess/P116/ft1/gll_ph_p116_r0239559565_v116.fit
/glast/Data/Flight/Reprocess/P116/ls1/gll_ev_p116_r0239559565_v116.fit
```

DataCatalog location and naming

Logical directory and group template:

```
Data/Flight/Reprocess/<reprocessName>:<dataType>
```

Note that the <dataType> field (following the colon) is a DataCatalog 'group' name, and file names are of the form r<run#>.

Naming examples:

```
Data/Flight/Reprocess/P116:FT1 r0239557414
Data/Flight/Reprocess/P116:LS1 r0239557414
```

Data Sample

The data sample for P116 reprocessing includes the following runs.

Block	First run MET	First run UTC	Last run MET	Last run UTC	#Runs	#Merit Evt	DataCatalog Source	Note
1	239557414	2008-08-04 15:43:34	242047683	2008-09-02 11:28:02	431	909,050,672	/Data/Test/Flight/Repro/ReproTest7	reprocessed for alignment fix
1	242053458	2008-09-02 13:04:17	307579060	2010-09-30 22:37:38	11413	24,906,606,023	/Data/Flight/Level1/LPA/	Standard Level 1 output (ignore StdIntent)
1 Subtotal					11844	25,815,656,695		
2	307585048	2010-10-01 00:17:26	311108362	2010-11-10 18:59:20	617	1,368,042,493	/Data/Flight/Level1/LPA/	Standard Level 1 output
3	311112219	2010-11-10 20:03:37	311669624	2010-11-17 06:53:42	98	215,382,591	/Data/Flight/Level1/LPA/	Standard Level 1 output changed to correct gtdiffsp parameters
Grand Total	239557414	2008-08-04 15:43:33	311669624	2010-11-17 06:53:42	12559	27,399,081,779		

Final number of selected photon events (in the FT1 and LS1 files) = 435,801,176

These data are destined for the FSSC and were made public as of 17 November 2010

These data (along with subsequent Level 1 data) are in the SLAC Astroserver as P6_public_v2.

Three extra runs were unintentionally processed as part of P116:

Run	Task Stream	Trigger	LAT configuration flag	Disposition
242429468	499	nomSciOps_trigTest	LAT_CONFIG = 1	retained
250687192	1948	hldVetoCalib_Hi	LAT_CONFIG = 0	removed
250692922	1949	hldVetoCalib_Lo	LAT_CONFIG = 0	removed

The first of these was deemed good for science, while the last two were not. Therefore, runs 250687192 and 250692922 have been retroactively removed from the datacatalog, astroserver and FSSC. (Removed from dataCatalog on 2/22/2011.)

Bookkeeping

- (This page): Define ingredients of reprocessing (processing code/configuration changes)
- Processing History database: <http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P116>
 - List of all reprocessings
 - List of all data runs reprocessed
 - Pointers to all input data files (-> dataCatalog)
 - Pointers to associated task processes (-> Pipeline II status)
- Data Catalog database: <http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp>
 - Lists of and pointers to all output data files
 - Meta data associated with each output data product

P116-FT1

(This task is roughly equivalent to the P105-FT1 task but with updates and modernization.)

Status chronology

- 11/29/2010 - Run 245403855 start time modified in datacatalog (1 second earlier), regenerate runFile.txt, and rollback stream 1019 to make astroserver happy. This adds three (3) events to that run. (Statistics above not updated)
- 11/17/2010 - reprocess final backfill block3 (98 runs)
- 11/12/2010 - discover that gtdiffsp params were incorrect (evtclass parm ignored, so all events got all diff rsp calc). Next (last) backfill block will be corrected.
- 11/11/2010 - gear up for reprocessing data block2 (617 runs)
- 10/18/2010 - 146 mergeClumps jobs ran out of CPU time. Changed to xsl batch queue and rolled back.
- 10/14/2010 - Validation complete, begin production for data block1
- 10/1/2010 - Begin task construction

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P116-FT1
---------------	--

Task Status	P116-FT1
Input Data Selection	MERIT
Input Run List	http://glast-ground.slac.stanford.edu/Decorator/Decorate/u38/Reprocess-tasks/P116-FT1/config/runFile.txt
Reprocessing Mode	reFT1
evtClassDefs	00-18-00
photon cut	pass6_FSW_cuts
eventClassifier	dataclean_classifier.py
eventClassMap	n/a
s/c data	FT2 from P105 (runs 239557414 - 271844560), then from current Level 1 production
ScienceTools	09-18-05 (SCons build)
Allowed Code Variants	redhat4-i686-32bit-gcc34, redhat4-x86_64-64bit-gcc34, redhat5-i686-32bit-gcc41, redhat5-x86_64-64bit-gcc41 (Optimized)
Diffuse Model	/afs/slac.stanford.edu/g/glast/ground/releases/analysisFiles/diffuse/v2/source_model_v02.xml (https://confluence.slac.stanford.edu/display/SCIGRPS/Diffuse+Model+for+Analysis+of+LAT+Data)
Diffuse Response IRFs	P6_V3_DIFFUSE, P6_V3_DATACLEAN
Output Data Products	FT1, LS1

Processing chain for FITS data products

Data Product	makeFT1	gtdiffrsp	gtmktime	gtltcube
FT1	true	true for evclass==3,4 (incorrect) evclsmin==3,4 (correct)	true	false
LS1	true	false	true	false

Note on 'Code Variant': The SLAC batch farm contains a mixture of architectures , both hardware (Intel/AMD 64-bit) and software (RedHat Enterprise Linux 4, and 5, gcc 3.4, 4.1, etc.). Currently, all batch machines have 64-bit hardware and (most, if not all) run with the 64-bit operating systems.

Note on diffuse response calculation: gtdiffrsp is called two times in succession. For the bulk of this processing, the incorrect parameter was used for gtdiffrsp (IRF P6_V3_DIFFUSE and evclass==3, and IRF P6_V3_dataclean and evclass==4). After block2 this was corrected to IRF P6_V3_DIFFUSE and evclsmin==3, and IRF P6_V3_DATACLEAN and evclsmin==4. The resulting FT1 files have five columns for diffuse response of which two pairs are filled in for DATACLEAN, one pair for DIFFUSE and zero for all other events. Each pair contains galactic and extragalactic response.

Timing

- The most time-consuming step in P116-FT1 is the 'mergeClump' step wherein the diffuse response is calculated (twice). The CPU time required for this step, after ~2300 jobs have completed appears in the following plot:

Thus, depending on the run and the speed of the batch machine (bimodal distribution for fell vs hequ), the jobs are finishing up after anywhere from 2 hours to 11 hours, with the most probable being about 5 hours.

With 1000 cores, one might then estimate 60 hours to complete the entire project. Near final timing

After the correct gtdiffrsp parameters were established, the mergeClumps step was reduced to ~30 minutes each.