

P300 - Full Pass8 Reprocessing

status: **Complete**

last update: 26 June 2014

This page is a record of the configuration and execution of the P300 reprocessing project, full reprocessing from DIGIs using Pass8 analysis code.



NOTICE

A [JIRA](#) was created for these tasks in May 2014. Look to this JIRA's comments for operational details that formerly were included on this page.

Pipeline tasks

- [P300-ROOT](#) - This [task](#) reads DIGI and produces reprocessed RECON + GCR + MERIT + FILTEREDMERIT (photons) + ELECTRONMERIT
- [P300-FITS](#) - This task will never happen – look for P301-FITS, which reads MERIT and produces FT1 (photons) + EXTENDEDFT1 + LS1 (merit-like FITS file for photons) + EXTENDEDLS1 + ELECTRONFITS file
- An experimental P300x-FITS [task](#) is being used for early validation studies. These data will not survive in the long run...
- L&EO data (264 runs) reprocessed with P300-LEO-ROOT and P300x-LEO-FITS

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v300.

XROOT location and file naming

Location template:

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

Locations for P300:

```
/glast/Data/Flight/Reprocess/P300/recon
/glast/Data/Flight/Reprocess/P300/gcr
/glast/Data/Flight/Reprocess/P300/relation/glast/Data/Flight/Reprocess/P300/merit
/glast/Data/Flight/Reprocess/P300/filteredmerit
/glast/Data/Flight/Reprocess/P300/electronmerit
/glast/Data/Flight/Reprocess/P300/ft1
/glast/Data/Flight/Reprocess/P300/extendedft1
/glast/Data/Flight/Reprocess/P300/electronft1
/glast/Data/Flight/Reprocess/P300/ls1
/glast/Data/Flight/Reprocess/P300/extendedls1
```

File naming:

Data Type	aka	Send to FSSC	Naming template
RECON		No	r<run#>_<version>_<dataType>.root
RELATION		No	r<run#>_<version>_<dataType>.root
GCR		No	r<run#>_<version>_<dataType>.root
MERIT		No	r<run#>_<version>_<dataType>.root
FILTEREDMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONFT1		No	gll_el_p<procVer>_r<run#>_<version>.fit
EXTENDEDFT1		No	gll_xp_p<procVer>_r<run#>_<version>.fit
FT1	LS-002	Yes	gll_ph_p<procVer>_r<run#>_<version>.fit
EXTENDEDLS1		No	gll_xe_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_FFD_RevA.pdf

Example:

```
/glast/Data/Flight/Reprocess/P300/recon/r0239557414_v202_recon.root
```

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/P300:RECON r0239557414
```

Data Sample

The currently defined data sample (as of 7/2/2013) for P300 reprocessing includes:

First run	239557414 (MET), 2008-08-04 15:43:34 (UTC)
Last run	present

Bookkeeping

1. (This page): Define ingredients of reprocessing (processing code/configuration changes)
2. Processing History database: <http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P300>
 - a. List of all reprocessings
 - b. List of all data runs reprocessed
 - c. Pointers to all input data files (-> dataCatalog)
 - d. Pointers to associated task processes (-> Pipeline II status)
3. Data Catalog database: <http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp>
 - a. Lists of and pointers to all output data files
 - b. Meta data associated with each output data product

P300-ROOT

Status chronology

- 4/23/2013 - Begin setting up P300 task for Pass8 reprocessing.
- 5/10/2013 - Initial version of task running. See this [page](#) for information about the task creation, and this [page](#) for performance comparison with P202
- 7/2/2013 - after weeks of testing, final GR, jobOpts, etc. come together and production tentatively starts.
 - **Block 1** is defined with this run range: 239557414 through 392842073 (2008-08-04 15:43:37 through 2013-06-13 18:47:50 UTC)
 - Update task to use the P202 generation of FT2 files (adds columns LAMBDA, RA_SUN, DEC_SUN, as well as meaningful BTIs)
 - First 10 trial jobs launched. Too many ACD INFO messages 'Missed Poca for ID 602 at...' so modify jobOpts to disable.
 - xroot disk space at the start of Pass8 = 508 TB (just after commissioning of fermi-xrd11)
- 7/29/2013 - a new GR is prepared (20-09-01) which fixes some problems in ACD tracking issues affecting ~32 runs so far (out of 5283). trickleStream disabled, wait for pending jobs to start, rollback other failed jobs, switch to new GR, rollback of ACD-failed jobs...
 - This new GR continues to seg fault, so revert back to GR 20-09-00 and restart trickleStream.
- 7/31/2013 - A new GR (20-09-02) with patch for ACD tracking installed after stream 5970. At the time, there were 33 failed processClumps jobs. All rolled back.
 - Last stream with GR 20-09-00 = 5970, run 274118943
 - First stream with GR 20-09-02 = 5971, run 274124672
 - Full list of rolled back substreams is [here](#)
- 10/15/2013 - Update job options to use the new "L1current" flavor of calibrations/alignment rather than "p7repro". First stream with new configuration: 20804.
- 11/14/2013 - Block 1 complete except for routine cleanup

- 11/21/2013 - **Block 1 complete**

Problems encountered include:

- ** pipeline_summary file corruption causing loss of email or of it being ignored, finally jobs terminate via reaper (Tony notified)
 - ** large blocks of failures due to dataCatalog not returning desired FT2 files (Brian notified, fix possible by moving query to scriptlet)
 - ** jobs running out the clock due to anomalously slow xroot behavior (Wilko/Andy notified)
 - ** problems with individual batch machines (Renata notified, and I also have the authority to remove batch machines from production)
- Summary of events processed in Block 1 from dataCatalog:

Name	Type	Files	Events	Size	Created (UTC)
ELECTRONMERIT	Group	26929	0	1.4 GB	09-May-2013 22:22:19
FILTEREDMERIT	Group	26929	0	1.4 GB	09-May-2013 22:22:17
GCR	Group	26929	58,653,424,708	1.0 TB	09-May-2013 22:22:18
MERIT	Group	26929	58,653,424,708	98.3 TB	09-May-2013 22:22:18
RECON	Group	26929	58,653,424,708	816.1 TB	09-May-2013 22:22:20
RELATION	Group	26929	58,653,424,708	7.7 TB	09-May-2013 22:22:20

- 4/30/2014 - Refactor P300-ROOT task to move all dataCatalog queries from batch jobs into scriptlets; update GPLtools (segregate INput and OUTput files); update git repository; etc.. Create task version 1.1 starting with stream 26929 to improve web app performance. Create JIRA for this task. Submit first 50 streams of 'block 2'
- 5/6/2014 - After big Oracle upgrade, regenerate/redefine 'block 2' to contain all runs through April 2014. This block contains 31838 runs, in increase of 4909 runs, starting with 392848044 and ending with 420590482. All future routine operational comments will be added to the JIRA and not to this confluence page.
- 6/8/2014 - Block 2 complete (see JIRA for operational notes)

Name	Type	Files	Events	Size	Created (UTC)
ELECTRONMERIT	Group	31838	0	1.7 GB	09-May-2013 22:22:19
FILTEREDMERIT	Group	31838	0	1.7 GB	09-May-2013 22:22:17
GCR	Group	31838	69,250,217,017	1.2 TB	09-May-2013 22:22:18
MERIT	Group	31838	69,250,217,017	115.9 TB	09-May-2013 22:22:18
RECON	Group	31838	69,250,217,017	961.1 TB	09-May-2013 22:22:20
RELATION	Group	31838	69,250,217,017	9.1 TB	09-May-2013 22:22:20

- 10/6/2014 - Setup Block 3, data after 30 April 2014 through 30 Sep 2014.
 - o Last run of Block 2 = 420590485 2014-04-30 22:41:22
 - o First run of Block 3 = 420596442 2014-05-01 00:20:39
 - o Last run of Block 3 = 433810379 2014-09-30 22:52:56
 - o Number of additional runs added by Block 3 = 2316
 - o Total number of P300 runs = 34154
 - o # streams in P300-ROOT (v1.1) = 4909
 - o Last stream in P300-ROOT (v1.1) = 31837

- 10/21/2014 - Block 3 complete

Name	Type	Files	Events	Size	Created (UTC)
ELECTRONMERIT	Group	34154	0	1.8 GB	09-May-2013 22:22:19
FILTEREDMERIT	Group	34154	0	1.8 GB	09-May-2013 22:22:17
GCR	Group	34154	74,250,808,780	1.3 TB	09-May-2013 22:22:18
MERIT	Group	34154	74,250,808,780	124.3 TB	09-May-2013 22:22:18
RECON	Group	34154	74,250,808,780	1.0 PB	09-May-2013 22:22:20
RELATION	Group	34154	74,250,808,780	9.7 TB	09-May-2013 22:22:20

- 2/2/2015 - Setup Block 4, data from 1 Oct 2014 through 31 Jan 2015
 - o Last run Block 3 = 433810379 2014-09-30 22:52:56
 - o First run Block 4 = 433816093 2014-10-01 00:28:10
 - o Last run Block 4 = 444436565 2015-01-31 22:36:02
 - o Total number of P300-ROOT runs = 36020
 - o Number of additional runs added by block 4 = 1866
- 2/17/2015 - Block 4 complete (after a week of disasters). Summary from dataCatalog:

Name	Files	Events	Size	Created (UTC)
GCR	36020	78,204,860,186	1.4 TB	09-May-2013 22:22:18

MERIT	36020	78,204,860,186	130.8 TB	09-May-2013 22:22:18
RECON	36020	78,204,860,186	1.1 PB	09-May-2013 22:22:20
RELATION	36020	78,204,860,186	10.2 TB	09-May-2013 22:22:20

- 4/8/2015 - Setup Block 5, data from 1 Feb 2015 through 7 Apr 2015 **penultimate** backfill
 - Last run Block 4 = 444436565 2015-01-31 22:36:02
 - First run Block 5 = 444442524 2015-02-01 00:15:21
 - Last run Block 5 = 450142738 2015-04-07 23:38:55
 - Total number of P300-ROOT runs = 37023
 - Number of additional runs in block 5 = 1003
- 4/16/2015 - Block 5 complete. Summary from dataCatalog:

Name	Files	Events	Size	Created (UTC)
GCR	37023	80,323,456,012	1.4 TB	09-May-2013 22:22:18
MERIT	37023	80,323,456,012	134.4 TB	09-May-2013 22:22:18
RECON	37023	80,323,456,012	1.1 PB	09-May-2013 22:22:20
RELATION	37023	80,323,456,012	10.5 TB	09-May-2013 22:22:20

- 6/4/2015 - Setup and start Block 6, through 6/3/2015
 - Last run of Block 5: 450142738 (2015-04-07 23:38:55)
 - First run of Block 6: 450148449 (2015-04-08 01:14:06)
 - Last run of Block 6: 455062807 (2015-06-03 22:20:04)

Total runs in combined runList = 37888 (formerly 37023)
Total new runs in Block 6 = 865
- 6/10/2015 - Block 6 complete. Summary from dataCatalog:

Name	Files	Events	Size	Created (UTC)
GCR	37888	82,148,825,606	1.4 TB	09-May-2013 22:22:18
RELATION	37888	82,148,825,606	10.8 TB	09-May-2013 22:22:20
MERIT	37888	82,148,825,606	137.4 TB	09-May-2013 22:22:18
RECON	37888	82,148,825,606	1.1 PB	09-May-2013 22:22:20

- 6/24/2015 - Setup and start Block 7, the **final** block of P300 reprocessing:
 - Last run of Block 6: 455062807 (2015-06-03 22:20:04)
 - First run of Block 7: 455069221 (2015-06-04 00:06:58)
 - Last run of Block 7: 456829490 (2015-06-24 09:04:47)
 - First run of Level 1: 456835199 (2015-06-24 10:39:56)
 - Total runs in full reprocessing task = 38198
 - Total new runs in Block 7 = 310
- 6/26/2015 - Block 7 – and entire task – COMPLETE

Name	Files	Events	Size	Created (UTC)
GCR	38198	82,812,280,309	1.5 TB	09-May-2013 22:22:18
MERIT	38198	82,812,280,309	138.5 TB	09-May-2013 22:22:18
RECON	38198	82,812,280,309	1.1 PB	09-May-2013 22:22:20
RELATION	38198	82,812,280,309	10.8 TB	09-May-2013 22:22:20

- 8/2/2015 Emergency reprocess of 40 runs that were Level 1 processed with bad CAL calibrations

Former (block7) run consisted of 38198 runs.
 New (block8) run consists of 38238 runs.
 There are 40 new runs to reprocess.
 First new run 459716687
 Last new run 459941302
 Starting point of pipeline task P300-ROOT (v1.1): 11269

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P300-ROOT
---------------	---

Task Version	1.0 (streams 0-26928), 1.1 (streams 26929-end)
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp
GlastRelease	20-09-00 streams 0 - 5970 (02 Jul 2013) 20-09-02 streams 5971 - end (31 Jul 2013)
Run Selection	based on a modified "standard" selection, see https://confluence.slac.stanford.edu/display/SCIGRPS/Official+LAT+Datasets (((sIntent=="nomSciOps" sIntent=="nomSO_noSk_noCno_optGccc_allEna" sIntent=="nomSciOps_diagEna" (sIntent=="nomSciOps_Emin5MeV"&&RunMin>242070455) nRun==242429468) && (RunQuality != "Bad" is_null (RunQuality))) sIntent=="nadirOps")
s/c data	P202 FT2SECONDS which will eventually become a "standard" Public Release https://confluence.slac.stanford.edu/display/SCIGRPS/Official+LAT+Datasets
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P300-ROOT/config/runList.txt
photonFilter	Not yet applicable, although this is defined by default: CTBParticleType==1 && ((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1 in evtClassDefs NOTE: task P301-MERIT regenerates MERIT files from P300 MERIT files using TMineExt, updating the event classification data
electronFilter	Not yet applicable, although this is defined by default: CTBParticleType==1
Code Variants	redhat5-i686-32bit-gcc41 (Optimized), note that rhel5-64 and rhel6-64 GR builds are not yet available
jobOpts	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P300-ROOT/config/Pass8Recon.txt
calibrations & alignments	"p7repro" and "L1current" (the latter to conform with new Level 1 flavor, and to include new ACD calibration for 9 Sep 2012)
Output Data Products	RECON , GCR , RELATION , MERIT , FILTEREDMERIT , ELECTRONMERIT (but note that FILTEREDMERIT and ELECTRONMERIT are empty or contain junk)

Timing and Scaling

- Current performance results for the Pass8 code are on this [page](#) .
- For 20,000 event clumps:

job step	average CPU time
processClump	183 min
mergeClumps	10 min

- Plot of job throughput as of 20130813
 .
 From the plot, one can see a processing rate of ~208 runs/day. Assuming 28500 total runs, this would mean a repro time of 137 days or 4.5 months.
- Another plot showing the number of run scratch directories cleaned up per day:
 .
- Timing plots for first part of task version 1.1

P300x-FITS

This task generates all desired FITS data products.

NOTE: a **temporary** task called P300x-FITS has been created. It generates only EXTENDEDFT1 files and nothing else. There is no diffuse calculation performed. An **untagged** version of evtClassDefs is used which contains Matthew Wood's initial event classification and selections. This will likely be a template for a future production P300-FITS task. All data produced by this task are "throw-away" and not expected to survive beyond initial validation studies.

Status chronology

- 8/1/2013 - Initial P300x-FITS task created. 10 trial runs processed. (See caveats above.)
- 8/7/2013 - First year of data run through the task, amounting to 5537 runs and 11790391606 events:

First run	239557417	2008-08-04 15:43:37
Last run	271850279	2009-08-13 09:57:59

- 8/28/2013 - Second year of data ready and processed -> extendedFT1. Note that 36 troublesome runs omitted from 1st year data were included in this sample. There should be no missing runs in this sample.

First run	239557417	2008-04-2008 15:43:37
Last run	302647722	2010-08-04 20:48:40

10,976 runs, 2,677,747,829 events, 234.7 GB

- 9/25/2013 - Third year of data ready.

First run	239557417	2008-04-2008 15:43:37
Last run	334184989	2011-08-04 21:09:47

This block contains 16,516 runs and 36,056,385,847 events, or an increase over year 2 of 5,540 runs.

- 11/21/2013 - Prepare for final Block 1 backfill. First run of backfill: 334190716, last run of backfill: 392842073, a total of 10 413 additional runs to process.
- 11/22/2013 - Block 1 is complete. The P300x area of the dataCatalog reports:

Name	Type	Files	Events	Size	Created (UTC)	Links
EXTENDEDFT1	Group	26929	7,149,992,671	626.6 GB	01-Aug-2013 17:29:46	Files

- THIS TASK IS DEFUNCT, replaced by P301-MERIT and P301-FITS

Configuration

CAUTION: The following data describes the *experimental* P300x-FITS task, not the production task

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P300x-FITS
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/exp/Fermi/task.jsp?task=112641219
Input Data	MERIT (direct from P300-ROOT)
spacecraft data	same as P300-ROOT
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P300x-FITS/config/runList.txt
evtClassDefs	untagged version from 20130801
eventClassMap	EvtClassDefs_P8.xml
ScienceTools	09-32-05 (7/30/2013)
Code Variants	redhat5-x86_64-64bit-gcc41 & redhat6-x86_64-64bit-gcc44 (Optimized)
Diffuse Model	based on contents of /afs/slac.stanford.edu/g/glast/ground/GLAST_EXT/diffuseModels/v2.0/v3r0 (see https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7)
Diffuse Response	N/A
IRFs	N/A
Output Data Products	EXTENDEDFT1

WARNING: THIS NEXT SECTION IS OBSOLETE

Generation of output data products:

Data Product	destination	data content [1]	event selection [1]	makeFT1	gtselect	gtdiffresp	gtmktime
EXTENDEDFT1	SLAC	FT1variables	((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1	✓	✗	✓	✓

FT1	FSSC+SLAC	FT1variables	'source' and above EVENT_CLASS bits 2,3,4 evclass=2 filtered from EXTENDEDFT1	✗	✓	(inherited)	✓
EXTENDEDLS1	SLAC	LS1variables	((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1	✓	✗	✓	✓
LS1	FSSC+SLAC	LS1variables	'transient' and above EVENT_CLASS bits 0,2,3,4 evclass=0 filtered from EXTENDEDLS1	✗	✓	(inherited)	✓
ELECTRONFT1	SLAC	FT1variables	CTBParticleType==1 pass7.6_Electrons_cuts_L1	✓	✗	✗	✓

[1] /afs/slac/g/glast/ground/releases/volume04/evtClassDefs/00-19-04/data

Note that diffuse response is calculated for 'source' and 'clean' event classes only.

Note on 'Code Variant': The SLAC batch farm contains a mixture of architectures , both hardware (Intel/AMD 64-bit) and software (RHEL5-64, gcc v4.1, etc.). At this time, GlastRelease builds only on RHEL5-32 (RHEL6-64 are built but not yet validated), while ScienceTools builds for RHEL5-64, RHEL5-64.

Timing and Scaling

(no data)

P300-LEO-ROOT

This task is a clone of the P300-ROOT task with the exception of the run list which contains the 264 L&EO runs.

12/10/2013 09:10 begin reprocessing

12/17/2013 Processing complete. Each of the RECON, MERIT, GCR and RELATION files contain 613,490,351 events.

P300x-LEO-FITS

This is a clone of the P300x-FITS task with the exception of the run list which contains the 264 L&EO runs.

12/17/2013 begin processing