P110 (Pass 7.2) Reprocessing Notes

P110 Reprocessing

status: **Complete**
last update: 15 April 2010

This page is a record of the configuration for the P110 reprocessing project, motivated by the Pass 7.2 event classification. This project involves reprocessing with Pass7 classification trees and (ultimately) new IRFs. The name "P110" derives from the word "processing" and the initial file version to be used for the output data products, e.g., r0123456789_v110_merit.root.

- P110-MERIT - this task reads DIGI+RECON+MERIT and produces reprocessed MERIT + FILTEREDMERIT (photons) + ELECTRONMERIT
- P110-FITS - this task will read FILTEREDMERIT and produce FT1 (photons) + LS1 (merit-like FITS file for photons) + electron FITS file + LS3 (live-time cube)

[Added 1 Feb 2010]

- P110-LEO-MERIT - like P110-MERIT but reprocess selected (earth limb pointed) L&EO data (see below for run list)
- P110-LEO-FT1 - like P110-FT1 but reprocess selected L&EO data

[Added 31 Mar 2010]
Added a one-week block of data around time of purported v407 Cyg X-1 flare (~11 Mar 2010)
Run range: 289873183-290564954

[Added 3 April 2010]
Added 231 runs from end of 11 Mar flare through early 2 April 2010
Run range: 290571022-291883927

[Added 15 April 2010]
Added 1 run corresponding to GRB 100414A
Run: 292903615

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v110.

XROOT location and file naming

Location template:

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

Locations for P110:

```
/glast/Data/Flight/Reprocess/P110/merit
/glast/Data/Flight/Reprocess/P110/filteredmerit
/glast/Data/Flight/Reprocess/P110/electronmerit
/glast/Data/Flight/Reprocess/P110/ft1
/glast/Data/Flight/Reprocess/P110/electronft1
/glast/Data/Flight/Reprocess/P110/ls1
/glast/Data/Flight/Reprocess/P110/ls3
```

For the P110-LEO data, the xroot locations are /glast/Data/Flight/Reprocess/P110-LEO/ etc.

File naming:

Data Type	Send to FSSC	Naming template
MERIT	No	r<run#>_<*version*>_<dataType>.root
FILTEREDMERIT	No	r<run#>_<*version*>_<dataType>.root
ELECTRONMERIT	No	r<run#>_<*version*>_<dataType>.root

ELECTRONFT1	No	r<run#>_<*version*>_<dataType>.fit
FT1	Yes	gll_ph_r<run#>_<version>.fit
LS1	Yes	gll_ev_r<run#>_<version>.fit
LS3	Maybe	gll_lt_r<run#>_<version>.fit

Example:

```
/glast/Data/Flight/Reprocess/P110/merit/r0239557414_v110_merit.root
/glast/Data/Flight/Reprocess/P110/filteredmerit/r0239557414_v110_filteredmerit.root
/glast/Data/Flight/Reprocess/P110/electronmerit/r0239557414_v110_electronmerit.root
/glast/Data/Flight/Reprocess/P110/ft1/gll_ph_r0239559565_v110.fit
/glast/Data/Flight/Reprocess/P110/electronft1/r0239557414_v110_electronft1.fit
/glast/Data/Flight/Reprocess/P110/ls1/gll_ev_r0239559565_v110.fit
/glast/Data/Flight/Reprocess/P110/ls3/gll_lt_r0239559565_v110.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.

Logical directories for P110:

```
Data/Flight/Reprocess/P110:MERIT
Data/Flight/Reprocess/P110:FILTEREDMERIT
Data/Flight/Reprocess/P110:ELECTRONMERIT
Data/Flight/Reprocess/P110:FT1
Data/Flight/Reprocess/P110:ELECTRONFT1
Data/Flight/Reprocess/P110:LS1
Data/Flight/Reprocess/P110:LS3
```

For the P110-LEO data, the DataCatalog locations are /Data/Flight/Reprocess/P110-LEO: etc.

In the DataCatalog, all file names are of the form r<run#>.

Naming examples:

```
Data/Flight/Reprocess/P110:MERIT r0239557414
Data/Flight/Reprocess/P110:FILTEREDMERIT r0239557414
Data/Flight/Reprocess/P110:FT1 r0239557414
Data/Flight/Reprocess/P110:LS1 r0239557414
Data/Flight/Reprocess/P110:LS3 r0239557414
```

Data Sample

The currently defined data sample for P110 and P110-LEO reprocessing includes:

	P110 (MET)	**P110 (UTC)**	**P110-LEO (MET)**	**P110-LEO (UTC)**
First run	239557414	2008-08-04 15:43:34	237928185	2008-07-16 19:09:45
Last run	277596392	2009-10-18 22:06:32	244406327	2008-09-29 18:38:47
Total runs	6581		199	
Total MERIT events	14,116,008,588		484,421,935	
Total FT1 events	2,358,821,051		138,013,907	

Note that the L&EO data represent a discontiguous set of runs.

Special one-week block of data around 11 March 2010

	P110 (MET)	P110 (UTC)
First run	289873183	2010-03-10 00:19:43
Last run	290564954	2010-03-18 00:29:14
Total runs	124	
Total MERIT events	272,014,125	
Total FT1 events		

Special block of 231 runs after March flare:

	P110 (MET)	P110 (UTC)
First run	290571022	2010-03-18 02:10:22
Last run	291883927	2010-04-02 06:52:07
Total runs	231	
Total MERIT events	509,989,399	
Total FT1 events		

Special run containing GRB100414A:

	P110 (MET)	P110 (UTC)
First run	292903615	2010-04-14 02:06:55
Total MERIT events	**3,156,688**	
Total FT1 events	**712,931**	

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=P110
    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

P110-MERIT

Status chronology

- 14 Apr 2010 - Processed 1 run for GRB (Nicola's request)
- 03 Apr 2010 - Processed block of 231 runs after March flare to 02 Apr 2010 (Richard's request)
- 31 Mar 2010 - Begin processing special block of 124 recent runs. Config is identical with the 2-run reprocess described in following bullet. Complete by 1 Apr 2010.
- 29 Mar 2010 - Rerun two runs (streams) from Oct 2008 which contain newly recovered data:

Run	UTC	Pipeline Stream	Previous # Events	New # Events
245403855	2008-10-11 07:44:15	1018	12,287	283,790
245409864	2008-10-11 09:24:24	1019	19,587	271,263

  For these two runs, the version of GlastRelease was updated from v17r35p1 to v17r35p1gr02, and the FT2 files were extracted from the P105-FT2 repository.
- 01 Nov 2009 - Processing complete
- 23 Oct 2009 - Xroot meltdown. Must meter jobs at ~600-800
- 22 Oct 2009 - Begin reprocessing remaining data (through 18 Oct 2009)
- 20 Oct 2009 - 650 early runs reprocessed (about 6 weeks, including two significant GRBs) with P110-MERIT

	MET(sec)	UTC
first run	239557414	2008-08-04 15:43:34
last run	243289793	2008-09-16 20:29:53

- 17 Oct 2009 - Single run reprocessed for validation

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P110-MERIT
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp
GlastRelease	v17r31p1
Input Data Selection	"standard" from https://confluence.slac.stanford.edu/display/SCIGRPS/LAT+Dataset+Definitions along with "&& (RunQuality != "Bad" || is_null (RunQuality)"
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P110-MERIT/config/runFile.txt
photonFilter	~~evtClassDefs~~ ~~v0r6p1~~ CTBParticleType==0 && CTBClassLevel>0
electronFilter	CTBParticleType==1
jobOpts	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P110-MERIT/config/reClassify.txt
Output Data Products	MERIT, FILTEREDMERIT, ELECTRONMERIT

Timing

P110-MERIT

The 650 runs in the six-week sample completed in about 20 hours elapsed time. Each run produces, on average, 7.5 1-hour "processClumps" jobs. Hence, the total CPU time to reprocess 650 runs is about 650 x 7.5 x 1 CPU-hour (fell-class machine) = 4875 CPU hours or 203 CPU-days.

The entire dataset (through 18 October 2009) consists of 6581 runs, which would be 49k CPU-hours or 2056 CPU-days. With 500 cores, this could take (with no operational problems) as little as 4.1 days.

P110-FT1

Status chronology

- 14 Apr 2010 - Added 1 run for GRB (Nicola's request)
- 04 Apr 2010 - Added 231 runs after March flare to 2 Apr 2010 (Richard's request)
- 01 Apr 2010 - Added 124 runs covering March flare (see above)
- 31 Mar 2010 - Re-reprocessed two runs to recover lost events (see above)
- 20 Nov 2009 - Processing complete
- 19 Nov 2009 - 12 of 6581 jobs require xxl queue to complete (due to enhanced fraction of diffuse photons - possibly due to ARR causing more albedo gammas - and to running gtdiffrsp three times)
- 18 Nov 2009 - All 6581 jobs complete, but with 287 time exceeded failures
- 17 Nov 2009 - 14:30 Begin production
- 16 Nov 2009 - Task configured, first test runs complete

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P110-FT1
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp
Input Data Selection	MERIT (from P110-MERIT), FT2 (from P100-FT2 and Level1)
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P110-FT1/config/runFile.txt
evtClassDefs	00-16-00
meritFilter	pass7_FSW_cuts, (FswGamState==0||FswGamState==3) && (CTBCORE>0) && (CTBBestEnergyProb>0) && (CTBBestEnergy>10) && (CTBBestEnergyRatio<5) && (CTBClassLevel>0)
eventClassifier	Pass7_Classifier.py
ScienceTools	09-15-05 (SCons build)
Code Variant	forced to redhat4-i686-32bit-gcc34

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	P7_v2_diff, P7_v2_extrad, P7_v2_datac
IRFs	implemented as 'custom irf', files in /afs/slac.stanford.edu/g/glast/ground/PipelineConfig/IRFS/Pass7.2
Output Data Products	FT1, LS1, LS3, ELECTRONFT1

Processing chain for FITS data products

Data Product	makeFT1	gtdiffrsp	gtmktime	gtltcube
FT1	true	true for evclsmin==8,9,10	true	false
LS1	true	false	true	false
LS3	false	false	false	true
ELECTRONFT1	true	false	true	false

Note on 'Code Variant': The SLAC batch farm contains a mixture of architectures , both hardware (Intel/AMD and 32-/64-bit) and software (RedHat Enterprise Linux 3, 4 and 5, gcc 3.2, 3.4, 4.1, etc.). GLAST/Fermi code builds on many newer combinations, but is not yet validated on them.

Note on diffuse response calculation: gtdiffrsp is called three times in succession. The first time with IRF P7_v2_diff and evclsmin==8, followed by IRF P7_v2_extrad and evclsmin==9, and finally IRF P7_v2_datac and evclsmin==10. The resulting FT1 file has six columns of diffuse response, two columns (galactic and extragalactic response) for each of the three IRFs. This creates a non-standard FT1 file by FSSC standards as they expect only five diffuse response columns.

Timing

The main batch job (mergeClumps) took <80 fell-minutes for the bulk of runs, but >24 hours for the last dozen or so runs.

P110-LEO-MERIT and P110-LEO-FT1

Configuration

The configuration for the "LEO" version of the reprocessing is mostly the same as for the ordinary science data with three exceptions: GlastRelease-v17r35p1gr02; the run list was provided by Anders (and consists of a discontiguous set of runs); and, the algorithm for finding FT2 files was modified to accommodate these earlier data (in fact, Warren produced a new set of 1-second FT2 files specifically for this reprocessing project). The run list for this reprocessing can be inferred from the list of merit files read by P110-LEO-MERIT, /nfs/farm/g/glast/u38/Reprocess-tasks/P110-LEO-MERIT/config/merit.txt. Note that the original list of runs counted 200, but a single run proved troublesome, 238781852, and was removed from the list, leaving 199 runs to reprocess.

Status chronology

- 01 Feb 2010 - Set up new tasks for L&EO reprocessing. These tasks behave like the original P110 tasks except that all output data products are stored in different directories both in xroot and the dataCatalog. Simply replace "P110" with "P110-LEO" to access these data.
- 07 Mar 2010 - P110-LEO-MERIT complete
- 10 Mar 2010 - P110-LEO-FT1 complete