

LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note LIGO-T010150-00 -E 11/30/01

**Naming Convention for Frame Files
which are to be Processed by LDAS**

P. Shawhan, S. Anderson, K. Blackburn, P. Ehrens, A. Ivanov,
A. Lazzarini, B. Mours, D. Sigg, J. Zweizig

Distribution of this draft:

All

This is an internal working note
of the LIGO Project

California Institute of Technology
LIGO Project - MS 51-33
Pasadena CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project - MS 20B-145
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

WWW: <http://www.ligo.caltech.edu/>

1 INTRODUCTION

This document establishes a naming convention for frame files which are to be processed by LDAS. The purpose of this naming convention is to allow LDAS to determine basic information about the contents of frame files from their names, without having to open and read the files, thus allowing an index to be constructed in an efficient manner. (The frameAPI currently maintains an index of frame files on disk, but the diskCacheAPI will take over this job.) Thus, given a time interval and detector specification, the software can quickly determine which frame file(s) contain the data for that time interval.

Frame files may be divided into four classes, depending on how they are produced and how they are to be used by LDAS:

1. “Raw” data files produced by the data acquisition system. These, naturally, have a rigid naming scheme which makes it straightforward for the diskCacheAPI to index them.
2. Processed, simulated, or otherwise constructed frame files in the data archive and/or LDAS disk cache. In general, a dataset of this type consists of a large number of files covering an extended time period, so these files too need to be indexed efficiently.
3. “Custom” datasets consisting of sequences of related files stored outside of LDAS. These can be processed by LDAS by specifying them explicitly by name in the LDAS user command; it is up to the user to determine which files are needed for any given LDAS job. Because these files do not need to be indexed by LDAS, there are different requirements on how they are named, but LDAS still needs to be able to infer time information from the filename in order to know which file (among those specified as inputs to a given job) contains the data for a given point in time.
4. Individual “custom” files which will be passed as inputs to LDAS jobs, but with only one input file per job. There is no ambiguity about which file to open to get the data, so there are no restrictions on the filename.

The rest of this document describes the naming convention for classes 1 through 3 above. Files in class 4 (as well as frame files that will never be processed by LDAS) can in principle have names of any form, though it will generally be a good idea to stick close to the naming convention described below even for these files. For example, everyone is encouraged to use the extension “.gwf” for all frame files, so that web browsers and Windows operating systems know how to treat them.

2 FRAME FILE NAMING CONVENTION

Frame files in classes 1 through 3 must have names of the form:

$$S-D-G-T.gwf$$

where:

- S* indicates the source of the data. For classes 1 and 2 (*i.e.* files which are to be indexed by LDAS), this consists of one or more uppercase letters in alphabetic order indicating the detector site(s), *e.g.* “G” for GEO600 or “HL” to indicate data from both Hanford and Liv-

ingston. (No distinction is made here between the two Hanford interferometers.) For class 3, this can be any string consisting of alphanumeric characters plus the underscore (`_`), plus (`+`), and number (`#`) symbols; the frame writer can choose to indicate a detector site, a user's initials, or something else in this field.

D is a description of the contents of a file. Files in class 1 (*i.e.* those written by the data acquisition system) will use a single uppercase letter, currently one of the following:

R full-rate “raw” time series data

T second-trend data

M minute-trend data

Files in classes 2 and 3 must have a description which **is not** a single uppercase letter, but can be any other string consisting of alphanumeric characters plus underscore (`_`), plus (`+`), and number (`#`). In particular, dots (`.`), dashes (`-`), and spaces are prohibited.

G is the GPS time (integer number of seconds) at the beginning of the first frame in the file. This is either a 9-digit or 10-digit number. (If the beginning of the data is **not** aligned with an exact GPS second, then the filename should contain the exact GPS second just **before** the beginning of the data.)

T is the total time interval covered by the file, in seconds. This field is required even if $T=1$. If the data is aligned with exact GPS seconds, then *T* is simply the number of seconds between the beginning of the first frame and the end of the last frame. (If the data is **not** aligned with exact GPS seconds, then *T* should be calculated from the exact GPS second just before the start of the data to the exact GPS second just **after** the end of the data. For example, a file which contains data from $t=700000035.42$ to $t=700000038.23$ would have a name like “H-R-700000035-4.gwf” even though the actual duration of the file is 2.81 seconds.)

Data “gaps” (*i.e.* non-contiguous frames within a file) are permitted, since ultimately the frameAPI will use the table-of-contents structure in the frame file to find the data it needs (or determine that it does not exist). However, as a matter of practice, we want the raw files produced by the framebuilder to **not** have gaps; if there is some missing data, the framebuilder should close the current output file and start a new one.

Note that this naming convention requires that a program which creates a frame file must ensure that the filename accurately represents the actual contents of the file. This may require the file to be renamed after it is written.

Summary of naming rules for each class:

Class 1: *S* indicates the detector site(s)

D is a single uppercase letter

Class 2: *S* indicates the detector site(s)

D can be anything **except** a single uppercase letter

Class 3: *S* can be anything

D can be anything **except** a single uppercase letter

3 EXAMPLES

Class 1

G-R-650123123-1.gwf	1 second of raw GEO data
H-R-688775627-8.gwf	8 seconds of raw LIGO data from Hanford
L-T-688775600-60.gwf	One minute of second-trend data from Livingston
HL-M-688775000-3600.gwf	One hour of minute-trend data from both Hanford and Livingston

Class 2

H-level2-693020060-256.gwf	256 seconds of “Level 2” data (<i>e.g.</i> selected channels, possibly with some offline processing) from Hanford
L-PEM-720878444-1000.gwf	1000 seconds of PEM data from Livingston
HL-SimInspiral-730282200-16.gwf	Some simulated data containing inspiral signals

Class 3

pss-filter_test-787384722-8.gwf	
iul-simset1-694215520-32.gwf	
H-calibPEM-720878444-1000.gwf	(The <i>S</i> field does not have to indicate the detector site, but it is permissible to do so.)