

LSC Detector Characterization
Working Group
Reduced Data Sets

Jim Brau

Gainesville

March 5, 1999

The goal is to reduce the data to a manageable volume.

As expressed by Keith Riles (Det. Char. Web Page):

There are two obvious ways:

1. eliminate extraneous channels
2. decimate the retained channels

The word reduction can have ambiguous meaning:

1. to make shorter
2. to transform the data (such as in "data reduction")

We will interpret the task as primarily 1., and some 2. may come in naturally

Primary task:

Define a data set which is significantly smaller than the full data set, is dynamic in content, which contains nearly a complete characterization of the detector, and the signals we are looking for.

Proposal:

Work with three types of data sets:

1. The full data set
2. a “standard” reduced data set
3. “designer” reduced data sets

The “standard” reduced data set will strive to be everything to everyone.

- It will include a complete set of the interesting channels, in a compressed format.
- It will be dynamic so that when it is found to be deficient, it can be revised.
- It will be designated “standard” by a designated LSC authority.
- It is in the light-weight data format

A “designer” reduced data set is created by a user who wants to make an alternative reduced data set.

- This allows one to test the completeness of the “standard”
- It is in the light-weight data format
- When the new elements of the “designer” set are found to be worthy of promotion, they may be added to the “standard” reduced data set
- Could also be used for short term purposes:
 - request two channels with no decimation for subtle correlation study
 - record a set of channels with one arm removed from the IFO via gross misalignment to study behavior of other arm
 - request full data set for a window triggered by satisfied set of conditions

We have in mind creating these reduced data sets by “eliminating” or “decimating” each of the channels of the full data set in a prescribed way, which can be channel dependent, with each channel reduced following one of a set of flexible, yet specific, algorithms.

For example, a set of such algorithms could be:

each channel of data is reduced according to one of a set of algorithms, including:

1. Full dynamic range data: no change in the data
parameters

A. data type: 1

2. Fourier transformed data
parameters:

A. data type: 2

B. bandwidth

C. data

3. Statistical data
parameters:

A. data type: 3

B. interval

C. mean

D. rms

E. peak

F. filtered peak

4. Flag data

parameters:

A. data type: 4

B. bound

C. flag (out of bounds, or within bounds)

5. Coarse binning

parameters:

A. data type: 5

B. max value

C. min value

D. no. of bits

6. Null: delete channel

parameters: NONE

7. Meta data

select type of metadata to preserve (perhaps all)

This suggestion has prompted a few comments already:

1. Trend data will be recorded by the DAQ in parallel with the ordinary ADC channels.
2. How does the standard reduced data set relate to the metadata database? For example, where should the out-of-bounds flags be stored? Does it make sense to store metadata in any data set, particularly the standard one?

Need to develop specific plan for implementing this, or another proposal

Keep it flexible!

Then, find volunteers to develop code to execute the plan

Assume the computing resources will be available at the observatories: CPUs and storage

Will want to be able to revise the standard reduced data set from the raw data at CACR....is this feasible?