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

Document Type	**LIGO-T030020-01-Z**	2003/06/09
S2 Playground Selection		
Lee Samuel Finn and Peter Saulson		

Distribution of this draft:

LIGO I Collaboration

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/

Processed with LaTeX on 2003/06/09

Abstract

We describe the playground selection criteria and final selections for the S2 run. The principle S2 playground data consists of a set of 600s samples of data choosen every 6370s throughout the course of the S2 run. In addition to this data, several longer segments are choosen randomly from triple coincidence time, several longer segments from double coincidence time involving at least L1, and representative 10 minute samples from the extremes of detector behavior based on DMT/glitchmon trigger rate.

$Id: T030020.tex,v 1.5 2003/06/09 23:36:10 lsf Exp $

Contents

1 Introduction

Cuts and thresholds for LIGO data analysis need to be evaluated on a set of data representative of that analyzed. The data set used for choosing these cuts is referred to as the *Playground* data.

The LIGO detectors are still in a state of rapid evolution; so, data from previous runs is not useful for the detailed choices that need to be made for the analysis of data from a particular run. Complicating matters further, LIGO detector data quality varies significantly over and between lock segments. This memo describes the procedure and final choices for the selection of Playground data for the S2 run, which attempts to meet the goal of providing a representative sample of data for the S2 data analyses approved at the beginning of the S2 run.

2 Playground components

The S2 Playground data consists of four components:

1. Every M seconds, starting at the beginning of the run and continuing to the end of the run, N seconds of data are set-aside from all three interferometers for the playground;

2. From the set of all triple-coincidence science mode lock segments longer than 2h in duration a total of three 30 minutes data segments are set-aside as playground;

Table 1: Maximum and minimum 10 minute averaged glitch rates in the three LIGO interferometers, and the corresponding GPS intervals set aside as playground. Glitch rates are from the 100 Hz band, $> 4\sigma$ channel of the dataqual dmt monitor.

Interferometer	H1	H2	L1
Max rate (Hz)	5.45	5.99	25.68
Playground interval	729730080–7297307680	731564100–731564700	734359740–734360340
Min rate (Hz)	0.0116667	0.0716667	
Playground interval	729695640–729696240	733074300–733074900	

3. From all double-coincidence science mode lock segments involving L1 and longer than 2h in duration, exactly three 30m segments are set-aside as playground;

4. From the science mode data in each IFO the 10m of data with the highest 10m average DMT/glitchmon trigger rate and the 10m of data with the lowest trigger rate.

Data selected by criterion 1 are, by definition, representative of the entire run. Should an analysis group desire to work with only the triple time, the triple time included in this subset will be representative of the whole triple time, etc. The choice of parameters M and N are discussed in section 3 below.

Data selected by criterion 2 and 3 give a few long, contiguous stretches to exercise on.

Data selected by criterion 4 provides examples of the "extremes" of detector performance, beyond that found in the representative samples in the subset selected by criterion 1.

Data selected by criterion 2, 3 and 4 are not determined until after the S2 run has completed. This document revision includes the preliminary specification of the data satisfying criterion 4.

2.1 Triple coincidence long-segment playground

2.2 Double coincidence long-segment playground

2.3 Extreme behavior playground

The extreme behavior playground data (cf. bullet 4 above) was identified using dataqual generated $> 4\sigma$ glitches on the AS_Q channel in the 100 Hz band. The maximum and minium rates and the 10 minute epochs set-aside for the playground are described in table 1.

3 Setting the sampling parameters M and N

3.1 Overview

In this section we discuss the sampling parameters M and N that determine how frequently data samples are set-aside for playground (M) and how long each playground sample is (N).

There are several constraints on the choice of M and N:

- The sample duration N must be sufficiently long that the analysis groups can use them for tuning their analysis. Tuning involves simulations and background estimation. To be

useful for simulations the playground segments should be many times longer than a signal duration and the minimum size data segment that is meaningfully used in the analysis. These constraints are determined by the analysis groups.

- The sampling period (M) should be incommensurate with "human" or "natural" timescales: e.g., the sampling period should not be commensurate with a a minute, hour, day or week period.
  - For this purpose it is useful to note that $86400 = 2^6 \times 3^3 \times 5^2$s and that many long-period anthropogenic noise sources tend to have start and stop times that align with hours ($3600 = 2^4 \times 3^2 \times 5^2$s), half-hours, quarter-hours, or even multiples of 10m periods within an hour.
- The ratio of sample duration to sampling period (N/M) should be in the range $7\% < N/M < 10\%$: i.e., no less than 7% and no more than 10% of the full data set should be playground.

3.2 Analysis group requirements

The four analysis groups (Bursts, Inspiral, Pulsar and Stochastic), the run manager (Whitcomb), the head of the detector characterization group (Riles) and the MIT lab head (Shoemaker) were all queried regarding the playground plans and any special playground data requirements they might have:

- Burst: time domain burst group analyses impose no significant constraints on the duration of a playground data segment. In S1 frequency domain analyses (TFCLUSTERS, POWER) worked with data in 360s segments. For S2 segments lengths will be longer up to an undetermined length, but not shorter.
- IULGroup: no response.
- Pulsar: the pulsar search currently estimates its noise on 60s intervals and plans to work with short Fourier transforms (SFTs) of 10m duration. Playground data segments should match this goal.
- Stochastic: no response.
- Run manager: no response.
- MIT lab head: commented that data segments (N) not be so short that our ability to detect longer-lived sources is compromised.

3.3 S2 choices

Bearing the above requirements in mind we choose

- the sampling period M to be 6370s ($2 \times 5 \times 7^2 \times 13$s; also 1h46m10s)

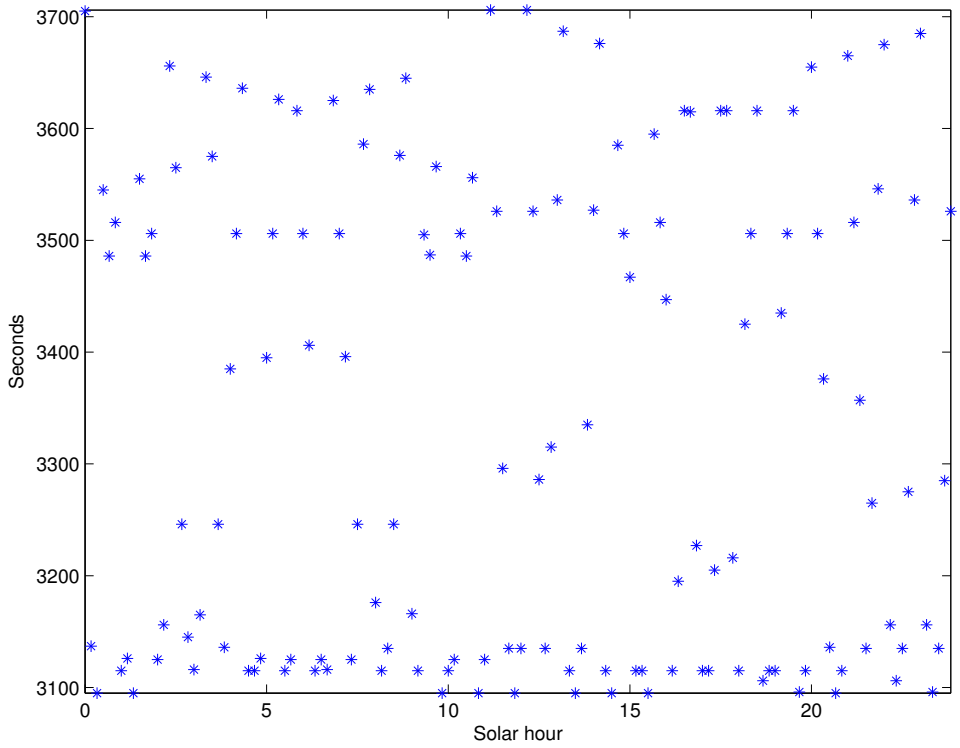

Figure 1: The number of seconds covered by the playground in each ten minute period of the solar day. The mean time spent in a ten minute segment beginning on the ten minute mark since the run beginning is 3389s and the minimum and maximum time spent is 3095s and 3706s.

- the sample duration N to be 600s ($2^3 \times 3 \times 5^2$s or 10 min).

The factorization of the sample period and the sample duration is chosen so that the the greatest common divisor between a natural period for anthropogenic noise and the sampling time is less than the sample duration. With these choices

- Playground constitutes 9.42% of the total run;
- The maximum length of an analysis data segment is 5770s;
- No second in the solar day is covered by less than 5 playground samples;
- No second in the solar day is covered by more than 7 playground samples;
- A sample begins in each solar hour twice every three days.

This choice of M and N distribute the sampling quite evenly throughout the 24h day. Figure 3.3 shows, the number of seconds sampled in the playground of each ten minute interval in a 24h solar day, while figure 3.3 shows the coverage, by second, of the playground during each day of the S2 run. In both figures time is measured from the beginning of the run and not from, e.g., 0h GMT.

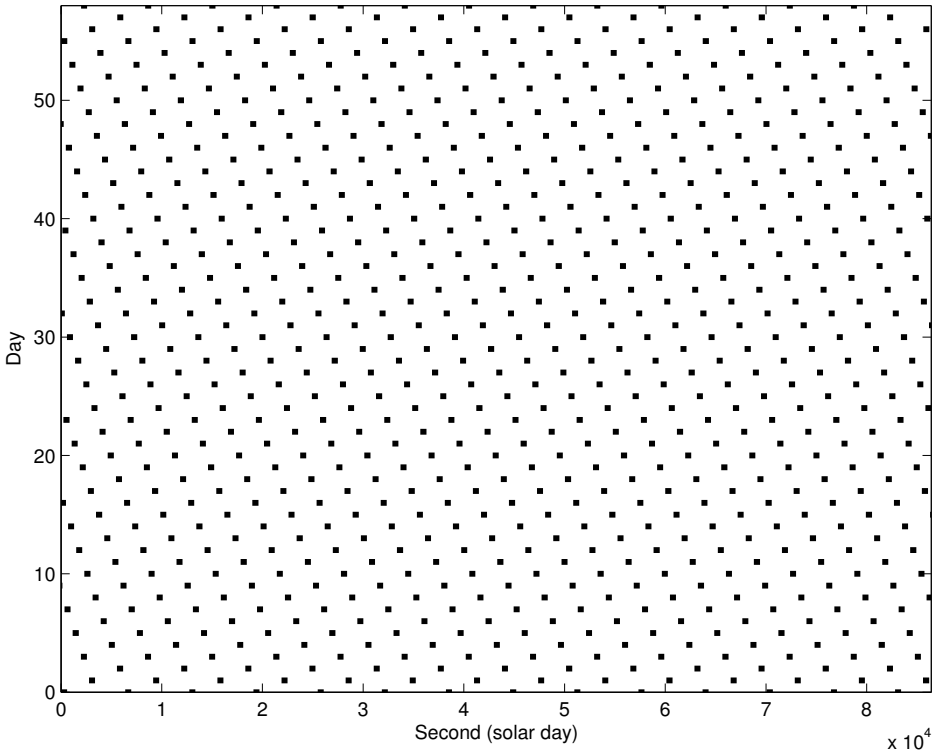

Figure 2: The coverage by the playground of each second of the solar day for each day of the S2 run.