A Common Data Format for Gravitational Waves Interferometers

Benoit MOURS
LAPP, BP 110, F-74941 Annecy-le-Vieux CEDEX, France

for the VIRGO Collaboration

ABSTRACT
The requirements for a data format suitable for gravitational wave interferometers will be discussed. A prototype of such a format with the associated software will be presented. The LIGO and VIRGO joint study on this topic will be reported.

1. Introduction

Several large interferometers designed for the detection of Gravitational Waves (GW) are under construction around the world. They are designed for permanent operation and will produce very large data set. They will eventually compare their signals and do coincidences searches. Therefore we need not only to design a data format suitable for our type of data but we should also prepare this future work. A minimal level of compatibility would be to use the same data architecture. A better one would be to to use the same data format even at the binary level.

In this paper, we first review the data format design issues, then we describe our selected solution and briefly discuss the current status and future developments.

2. Data Format Design Issues

Let us first remind that the interferometric detectors will provide a large amount of data mainly because a small number of channels are sampled at high frequency. This means that we can afford a little overhead for the data organization of all the channels, as long as we stay efficient for the high frequency channels. Therefore, the primarily idea is to group all the information necessary for the understanding of the interferometer behavior in units called frames. A frame contains high frequency and low frequency channels as shown by figure 1. The frame duration is not fixed a priori but will be typically of one or a few seconds. At this point we should notice that a GW event (binary coalescence, supernova,...) will last several frames.

The foreseen searches are of different types. There will be searches for burst events where we will need all the possible detector monitoring information for a few frames. There will be pulsar searches where we just want to integrate one signal but for a very long time. This implies that we need a flexible format where we can remove information without changing the top structure. We need also to be able to add information after the frame building done by the online data acquisition system. The information we want to add is for instance online trigger information, reconstructed h signal,... It could be also simulation information if the frames are produced by a

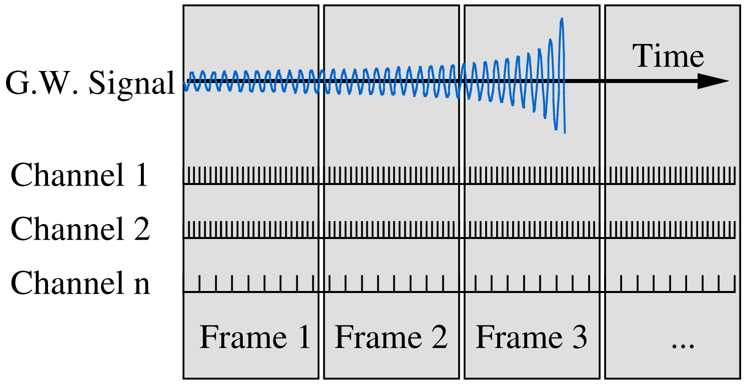

Figure 1: The frames are the organized blocs of data which include many samplings

Monte Carlo program. So, the flexibility of the format is one of the key issues and this will be achieved by using a tree like organization.

The format and its associated software (the frame library) has also to fulfill several technical requirements. Among them, let us mention that it should contain the description of the data stored to simplify to distribution of information and to provide some mechanism for format evolution. The input/output functions have to be efficient to handle the large data flow of the online system (several MBytes/s). It should the simple and portable on a wide type of platforms.

3. The Frame Organization

To implement the tree like structure and to simplify its manipulation, a frame is organized as a set of C structures described by a header holding pointers to additional structures and values of parameters expected to be stable over the integrated time interval: the starting time of the frame, its duration, values produced by the slow monitoring. This header is followed by an arbitrary number of additional structures, each holding the values of a rapidly varying parameter like the main signal, the seismic noise. Each active element producing data at a rate higher than the frame rate is thus accumulated in a dedicated structure.

This frame structure is a standard which has to be conserved over the various stages of the analysis. Frame history, detector geometry, trigger results, monitoring data, reconstructed data, simulation results lead thus just to additional structures. It is always possible to add new structures or to drop old ones. The figure 2 presents a schematic view of this organization.

4. The status

In 1996, a LIGO/VIRGO collaborative effort started to investigate the possibility of using a common format. It has recognized the need for a specific software which can handle efficiently our type of data. Starting from a software developed by Virgo, the two groups have identified the needed improvements and defined a set of tests. The software developments are still in progress and the format and associated software is being exercised on the LIGO 40m and on the VIRGO data acquisition prototype. These efforts should converge on a common software which will be publicly available.

In addition, an interface/translation to other public formats will be studied as a possible future path for public distribution of gravitational data "products" to complement the frame library software.

Acknowledgments

I would like to thank the LIGO team and especially K. Blackburn for stimulating discussions.

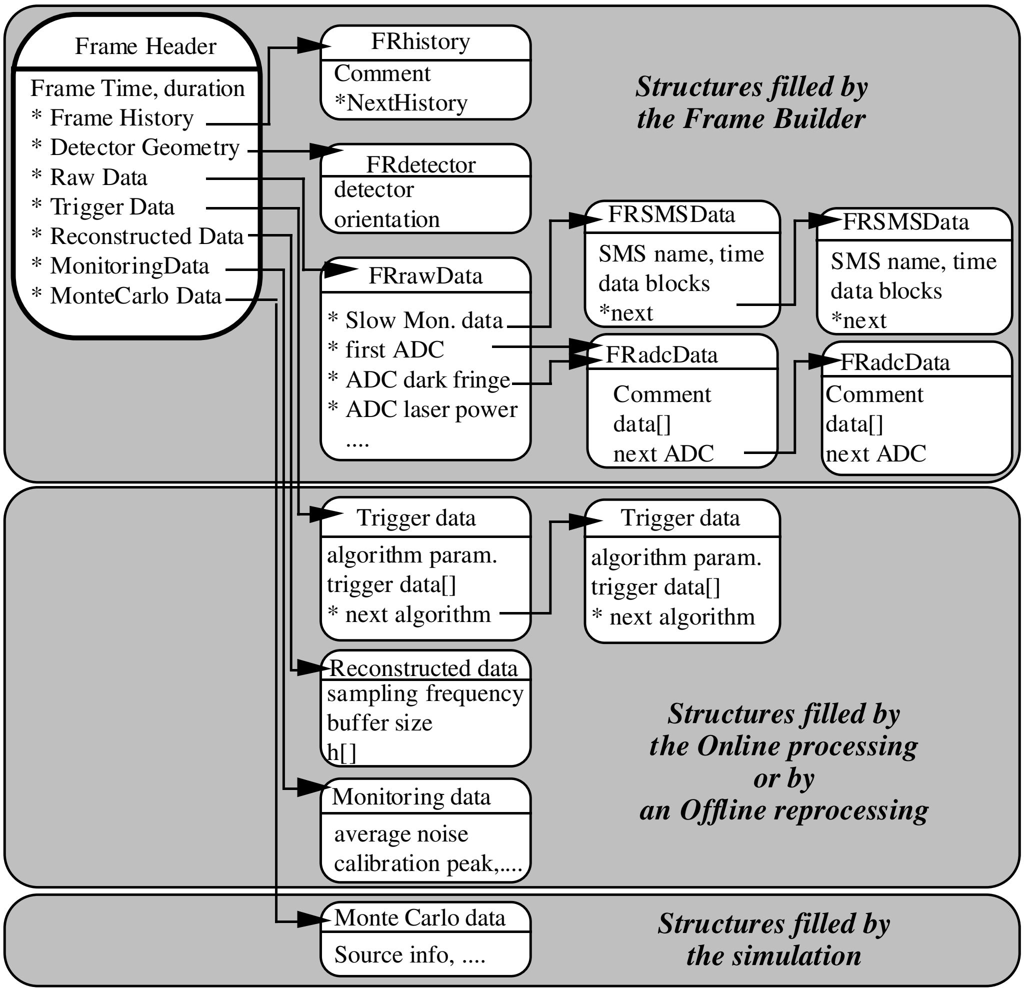

Figure 2: The tree like structure used to stored the data. Only a part of the structures and of their contents is shown in this figure.