

Grid Development and Deployment of LIGO Scientific Collaboration Data Analysis Software

Duncan A. Brown for the LIGO Scientific Collaboration

Onasys

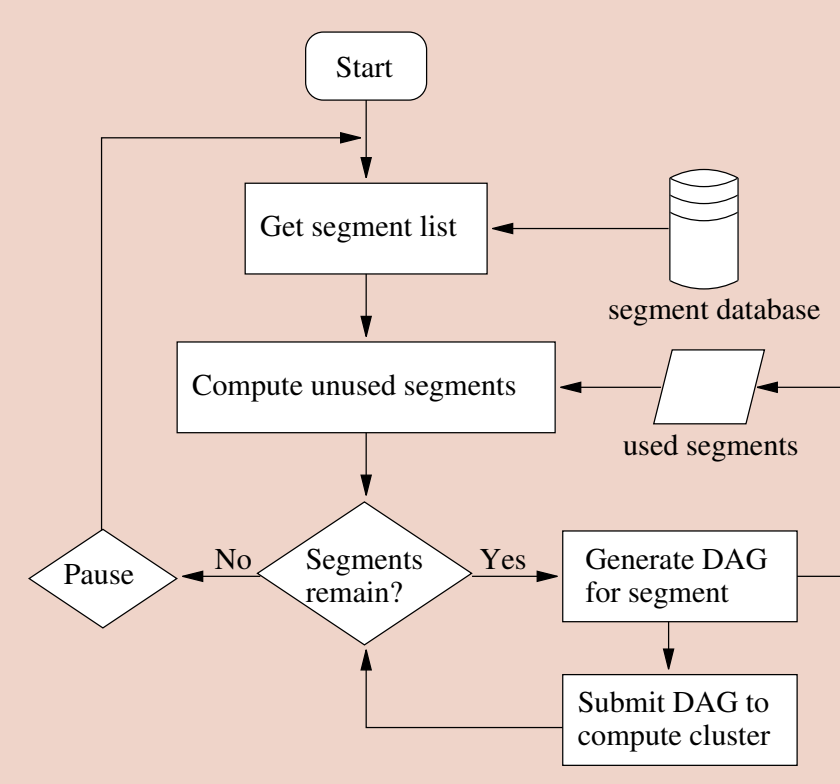
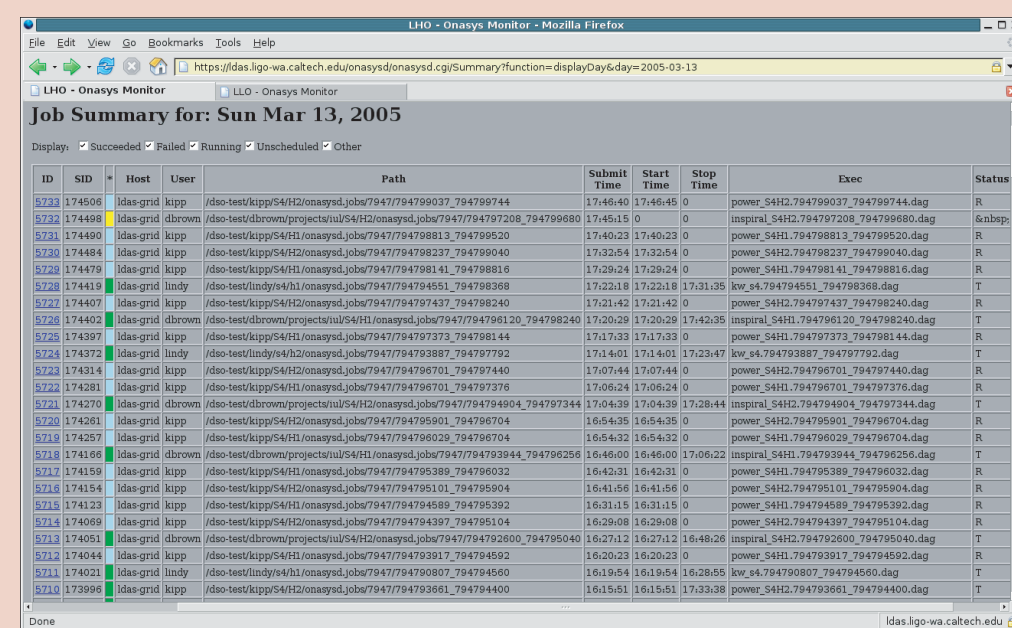
The ONline Analysis SYStem (onasys) software provides tools for automating the real-time analysis of gravitational wave data.

Executes data analysis pipelines created by LSC scientists in real-time to process data from the three LIGO gravitational wave interferometers as it is taken.

Responsible for the execution and monitoring of data analysis pipelines which generate astrophysical results. Results are returned to the user or can be published for replication around the LSC data grid.

Built on top of the Grid LSC User Environment (Glue) and uses the scientific data analysis pipeline software deployed by the LSC.

Online monitoring of analysis is available via a web interface which queries job information metadata databases.



Onasys periodically, and on a short time-scale, performs the following sequence: Identify the data to analyze using a query to a GSI-authenticating instrument status and data quality server (LSCsegFind).

Locate the data using a query to a GSI-authenticating data location server (LDRdataFind).

Configure a data analysis pipeline using a user-supplied pipeline construction tool.

Execute the data analysis pipeline on the grid.

LSC Users

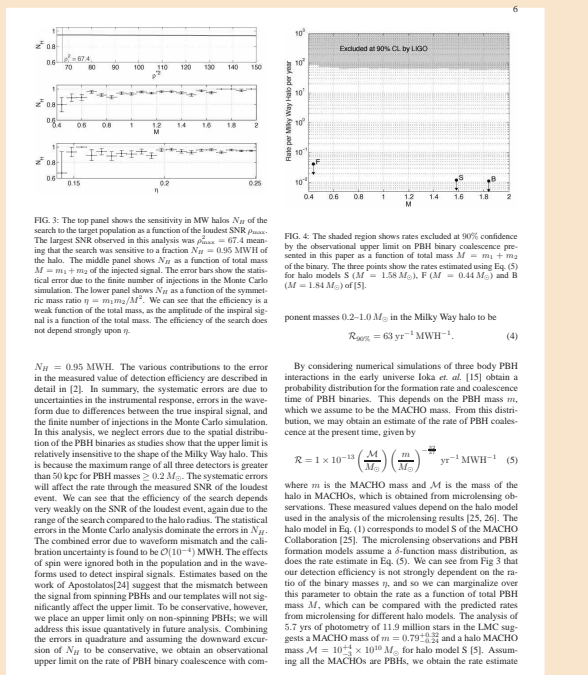
Users search for gravitational waves in LIGO data by running data analysis pipelines on the grid.

```
inspiral_pipe.ini (-/projects/lalapps/src/inspiral) - VIM
[defaults]
: use only level 1 rds data
type = RDS_R11

[Data]
: data conditioning parameters common to tsplitbank and inspiral
pad-data = 8
segment-length = 1048576
number-of-segments = 15
sample-rate = 4096
resample-filter = Ldas
resol-to-high-pass = 100.0
high-pass-order = 8
high-pass-attenuation = 0.1
spectrum-type = median
```

Simple text files contain the scientific parameters of the desired search.

Pipelines are run on the grid to produce scientific results



Pipeline Generation

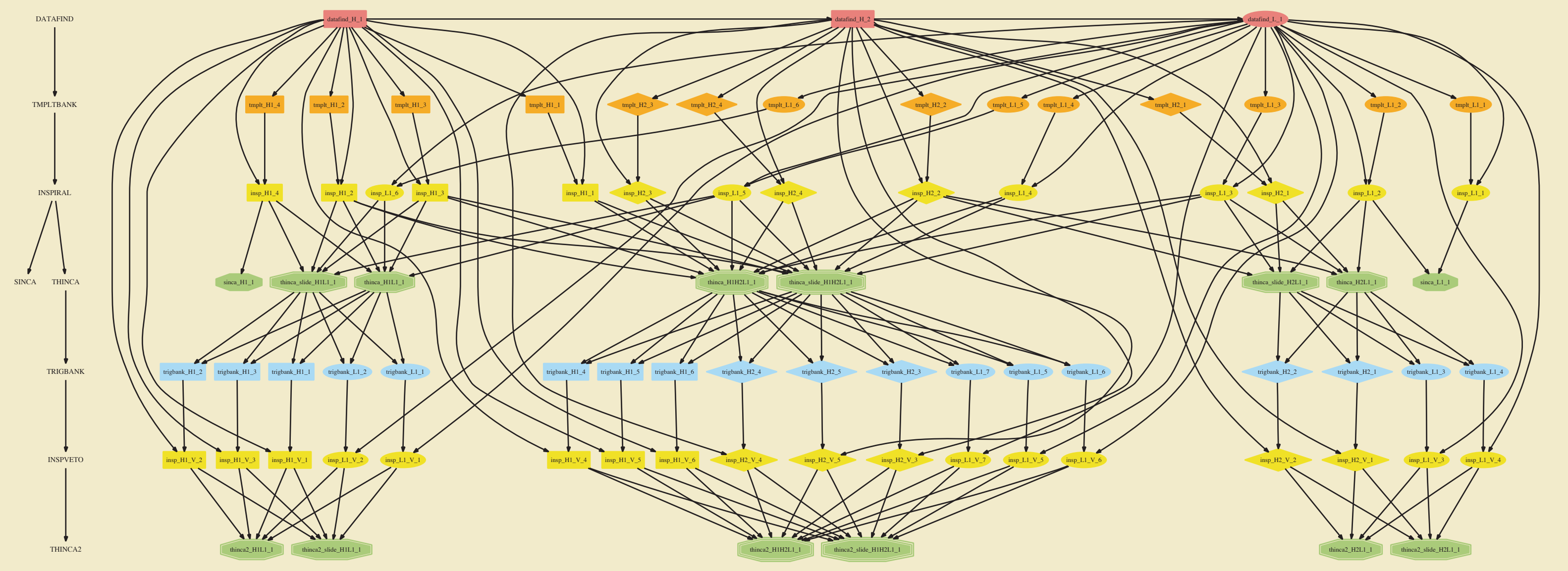
Complicated workflows can be constructed to perform all the steps necessary to search data from the four LSC detectors: the 3 LIGO interferometers and the GEO600 interferometer.

Pipeline generation is built on top of Glue and the LALApps analysis codes.

Different analysis strategies can be pursued by constructing various workflows.

Different pipelines are typically constructed for online analysis, offline follow-up of online triggers, and offline analysis used in large scale Monte Carlo simulation and parameter tuning.

Pipelines are created as Condor DAGs or VDS abstract workflows (DAX)



A subset of a binary inspiral workflow. Workflows used in analysis pipelines typically have over 10,000

Glue

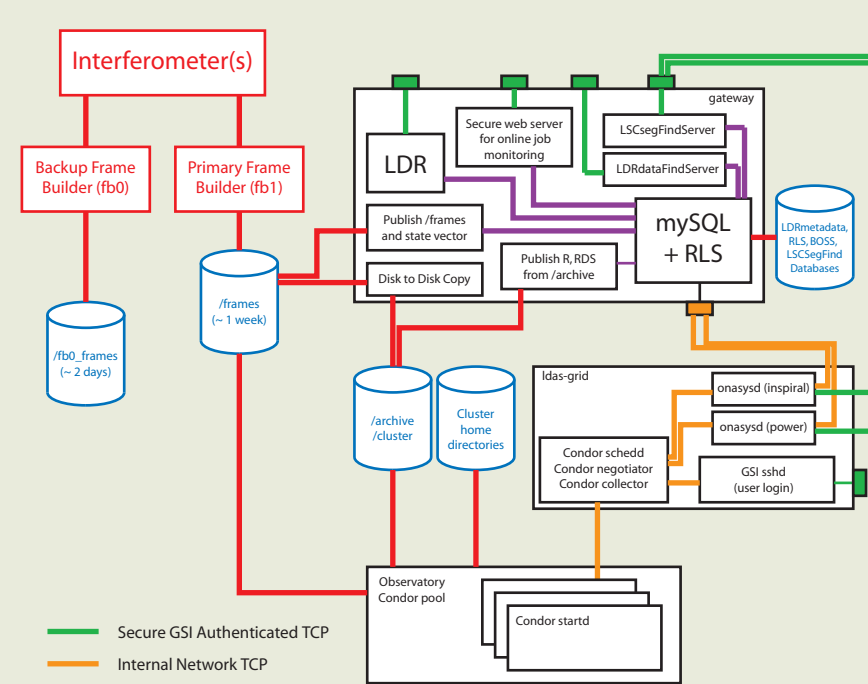
The Grid LSC User Environment (Glue) provides a high level infrastructure for developing LSC grid applications and metadata services.

Glue provides an infrastructure to simplify the construction of workflows by treating data analysis applications as modules that can be chained together.

Glue's use of metadata (e.g. data quality information) allows complicated workflows to be easily constructed.

Glue also contains certain LSC specific metadata clients and servers, such as data discovery tools.

The figure shows the online analysis infrastructure which uses components of Glue.



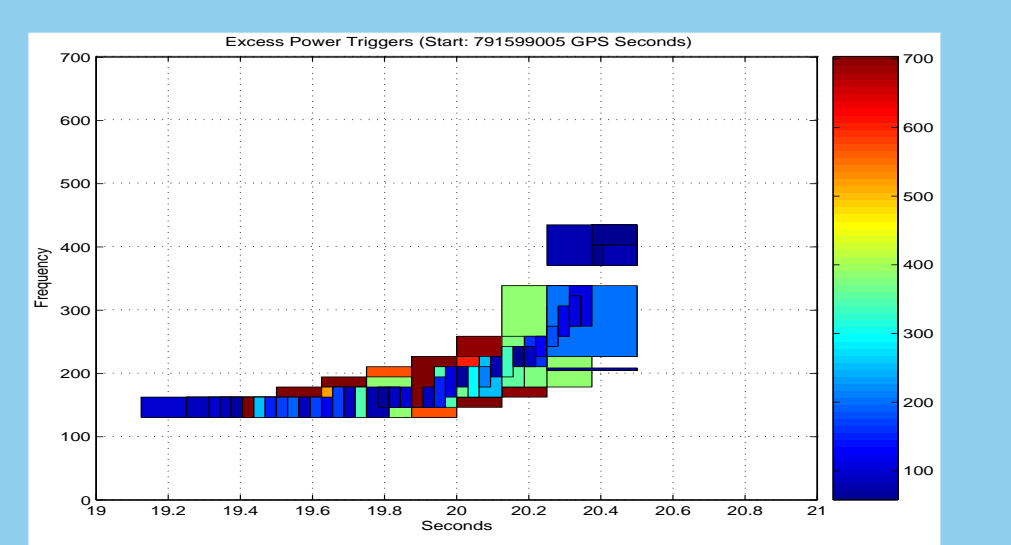
LALApps

User applications which perform specific data analysis tasks as part of a pipeline.

LALApps is a suite of stand-alone scientific data analysis applications built on top of LAL.

This suite contains programs for analyzing interferometer data as well as manipulation of data products.

It also contains code to abstract the scientific applications to allow pipeline construction in conjunction with Glue.



A simulated inspiral analyzed with the LALApps excess power search.

LDR

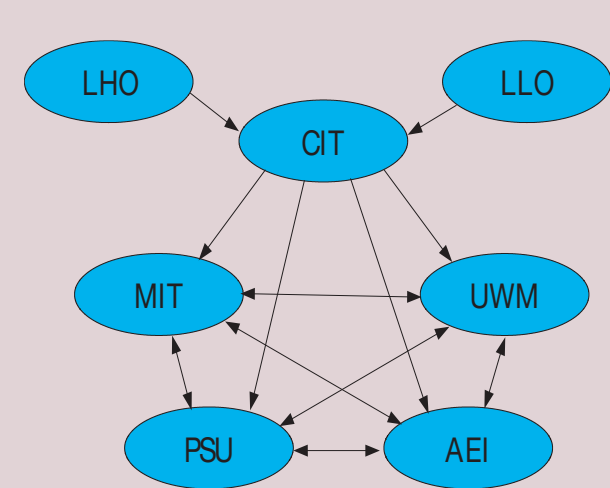
Replicates LIGO and GEO data around the LSC grid.

Maintains and distributes metadata about interferometer data.

Replicates data from LIGO observatories and between LSC compute sites.

Provides services to users and applications for mapping metadata to physical data locations.

Built on top of GridFTP, MySQL, Globus RLS and pyGlobus.



Pegasus

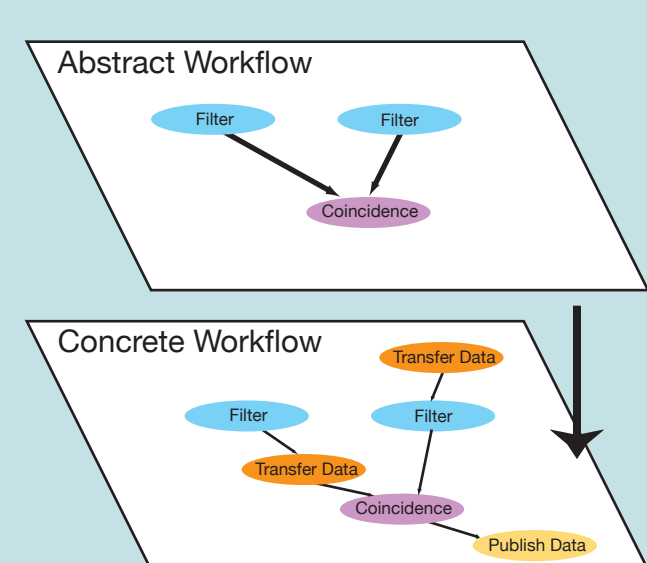
Allows analysis to run on grids which do not have LIGO data.

Takes abstract workflows created by data analysis pipeline generators.

Queries metadata catalogs to discover locations of data needed by each node in pipeline.

Adds additional data transfer and publication nodes needed to run pipeline on the grid.

Resulting concrete workflows can be executed on a wide variety of grids using Condor and Globus.



Condor and VDT

Provides grid middleware used to execute searches.

Condor provides the underlying infrastructure for execution and management of workflows on the grid, either directly or through submission to Globus job managers.

The LSC Data Grid client and server packages are built on top of the VDT. The client and server provide the subset of the VDT used by the LSC and add LSC specific grid tools. Pacman makes this bundling and distribution simple.

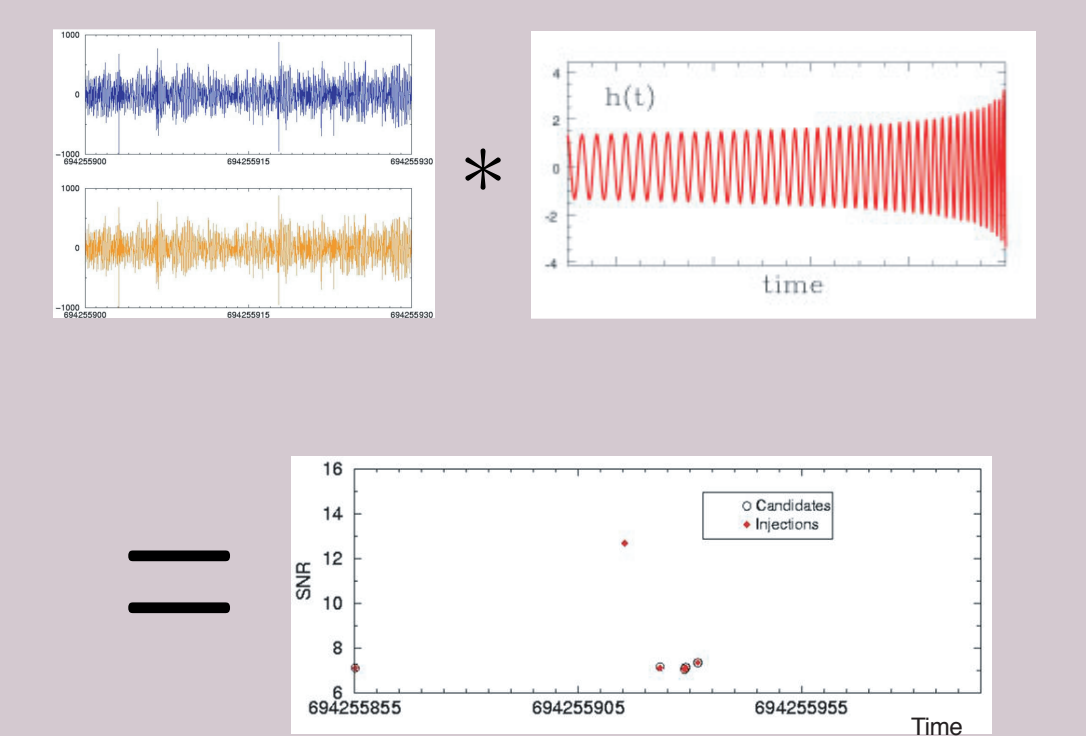
LAL

LSC Algorithm Library (LAL) contains data analysis routines.

Algorithms are written in ANSI C89 for wide portability. Routines are contributed to LAL by LSC members.

LAL algorithms include:

Data conditioning, gravitational wave simulation, correlation of data with binary inspiral signals, excess power filters, etc.



LSC Data Grid / Open Science Grid

Provides underlying hardware and operating system software for executing LSC data analysis code.

Computing clusters at Caltech, MIT, LIGO Livingston Observatory, LIGO Hanford Observatory, UWM, Penn State, AEI, Cardiff, University of Birmingham comprise the LSC Data Grid.

Online analysis is performed at the observatory sites (LHO and LLO). Offline analysis and follow-up of online results runs at other sites.



The LSC Data Grid provides a standard set of grid tools for users to execute pipelines.

Access to the LSC Data Grid is via ssh login or Globus job managers.

LSC Data Grid usage can be monitored using Ganglia.

Workflows constructed by Pegasus from the LSC workflows can be executed on the Open Science Grid.



OSG will be used for compute intensive workflows which are too large for the LSC grid.

