Omega Pipeline

status report

Jameson Rollins

LVC Meeting
September 20, 2009
Eötvös Loránd University
Budapest, Hungary

LIGO-G0900810-v2

Outline

- Overview
- Bayesian followup
- Online analysis
- New projects/extensions

Omega overview

- Matlab-based search pipeline
  - modular
  - simple development environment
- Next generation of Q/X-Pipeline
  - sine-Gaussian basis
  - Q-tiles tested for significance, clusters, coincidence
  - coherent event followups on significant tiles/clusters
- Currently provides important tools for the LVC:
  - Omega Scans – used by many groups
  - single IFO online trigger generation for glitch studies

Current Contributors

Ruxandra Bondarescu	Meagan Lang
Laura Cadonati	Nicolas Leroy
Eric Chassande-Mottin	Satyanarayan Mohapatra
Shourov Chatterji	Jameson Rollins
Mark Edwards	Antony Searle
Brennan Hughey	Patrick Sutton
Jonah Kanner	Leo Stein

What's new since S5Y2

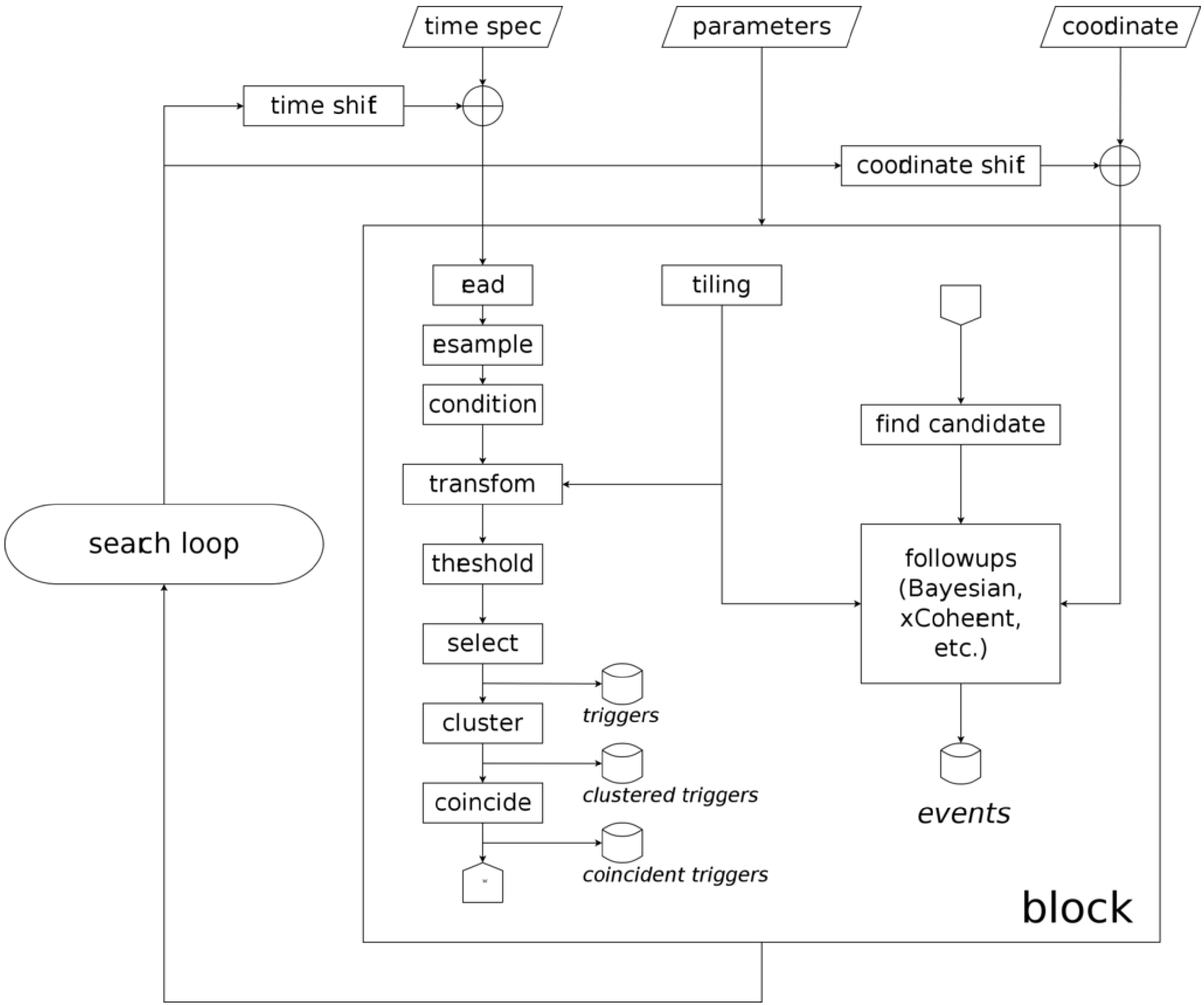
- Double whitening / matched filtering
  - previous versions searched for sine-Gaussians in whitened data, not for sine-Gaussians in physical *strain*
- Clustering
  - gather power from multiple nearby tiles
- **Bayesian followup**
  - coherent robust detection statistic
  - pointing with skymaps
  - interfaces to LAL Bayesian skymap code (also used by CBC)
- Sky-directed coherent searches
  - with both Bayesian and xCoherent followups

Other code changes

Much work has been done to make the pipeline more *reliable* and *debuggable*:

- More memory efficient
  - loops rearranged to eliminate temporaries
  - certain quantaties recomputed rather than stored
  - excess toolboxes removed
- Eliminated redundant code
- New logging, debugging, file outputs
- New post-processing
- Improved build system

Pipeline flowchart



Bayesian coherent followup

Motivation

Markov-Chain Monte Carlo analyses, as employed in CBC and LISA, are the gold standard for detection and parameter estimation because they explicitly check the data against all possible signals, effectively performing a computational inversion of the injection process.

However, this performance comes at a high computational cost; MCMC are impractical for the vast parameter space of all-sky burst searches.

Motivation

With Bayesian approach, we can use a conventional pipeline to crudely estimate some of the signal parameters, and then fully explore a subset of parameters of particular interest (such as direction).

The Bayesian statistic quite unlike any of the statistics previously used.

It's computationally challenging and hard to implement correctly, but with the potential for an impressive payoff in performance and conceptual simplicity.

Application

- Conventional pipeline produces candidate triggers
- Trigger frequency, Q, and approximate arrival time used as priors to reduce parameter space
- Marginalize over amplitude, phase, polarization and precise arrival time
- Subtract out proposed signal and compute the probability of the residual given the noise power
- Integrate this probability over all signal parameters except direction to get a “skymap” of directional probability

Application

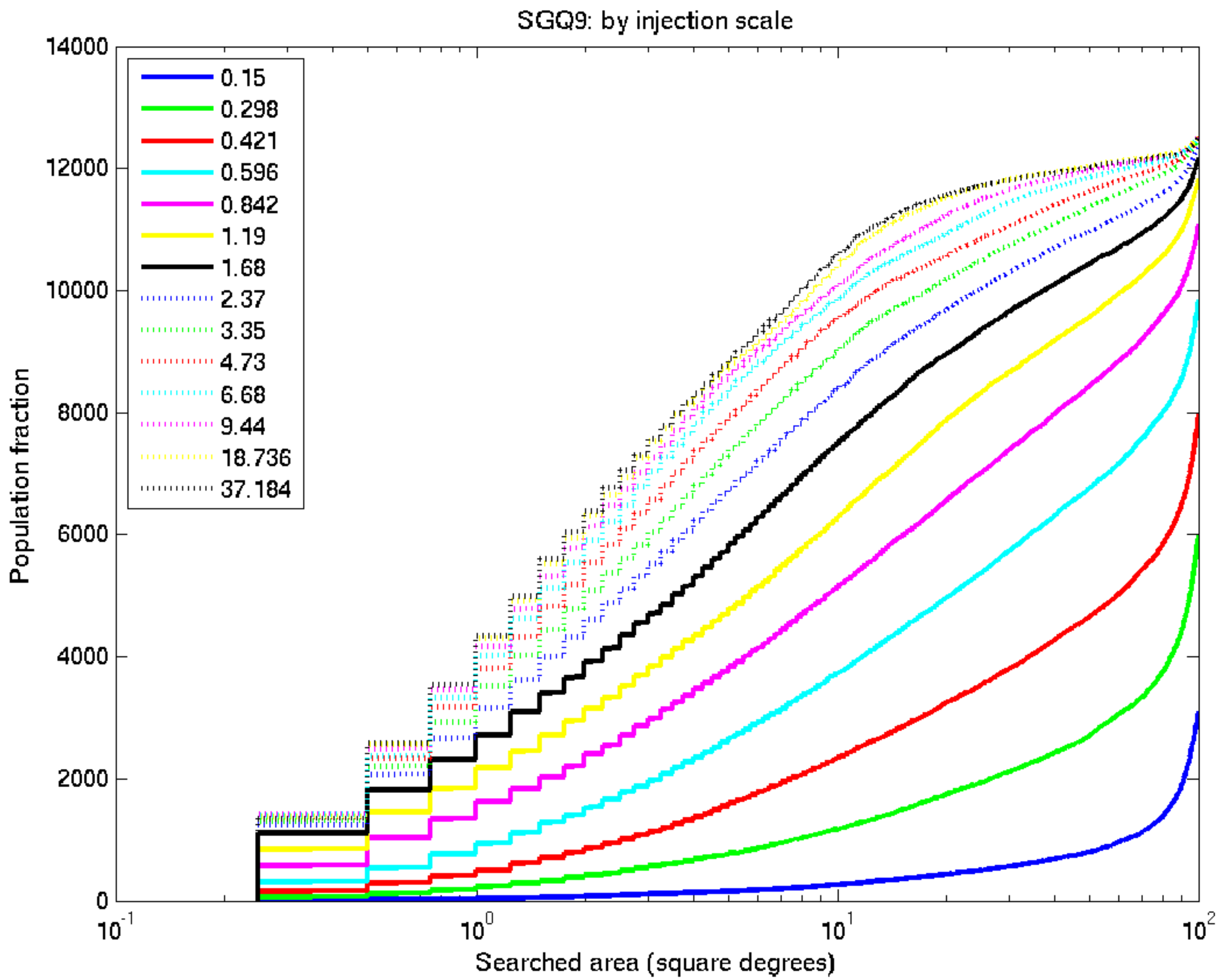
- Integrate over direction to get a scalar odds of signal vs. colored noise
- Produces the same statistic individually for each detector
- Comparison between the individual and coherent Bayesian odds gives us a Bayesian statistic capturing the coherence between the instruments

Status

- Everything is implemented
- Pointing works
  - time delays, antenna patterns, etc. are correct
- Position reconstruction is good

Position reconstruction

- Median asymptotes to two square degrees for large signals
- Skymaps are the right shape



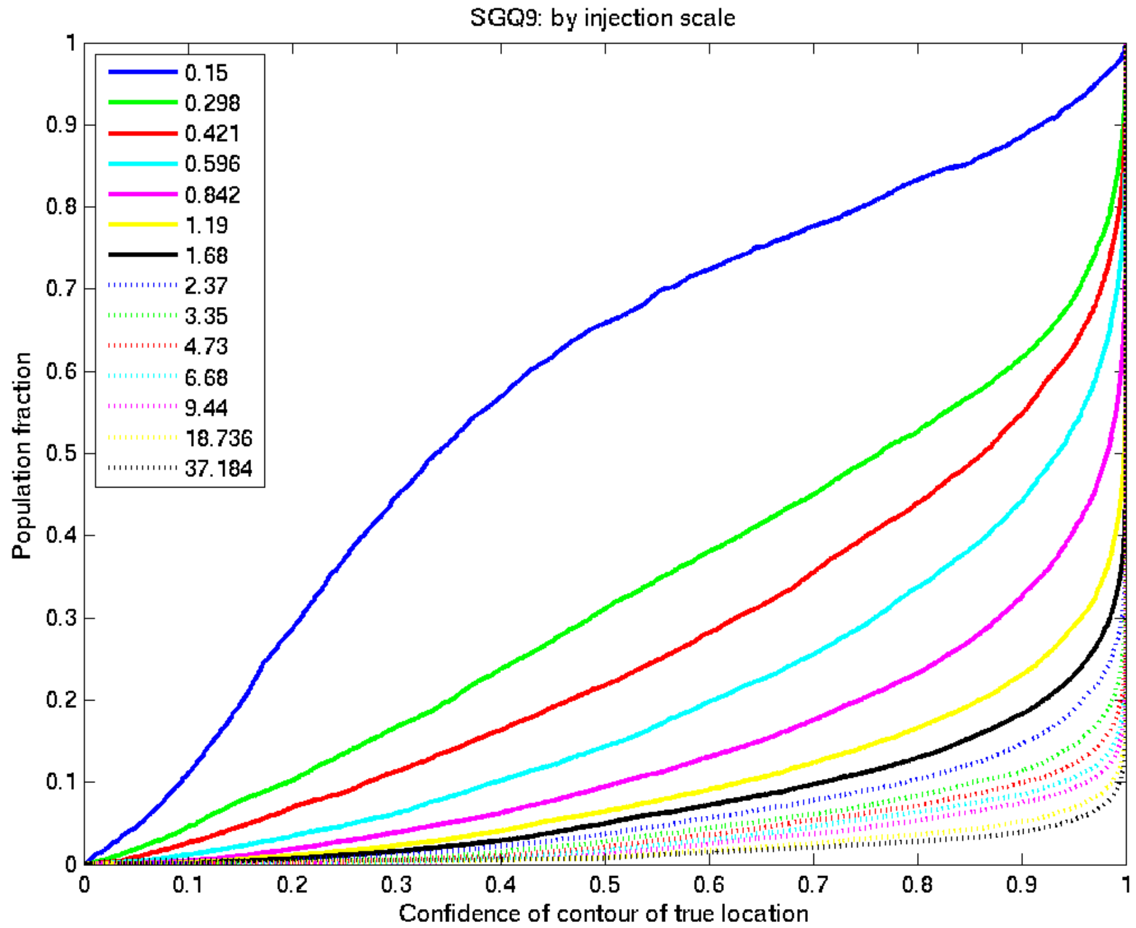
Issues

However, there are some issues...

Skymap calibration

Calibration of skymap contours is wrong

- Half of the injections should fall within 50% convidence contour region (ie. lines should be diagonal)
- Could use simulations to calibrate the contours as cWB does, but a symptom that something else is wrong



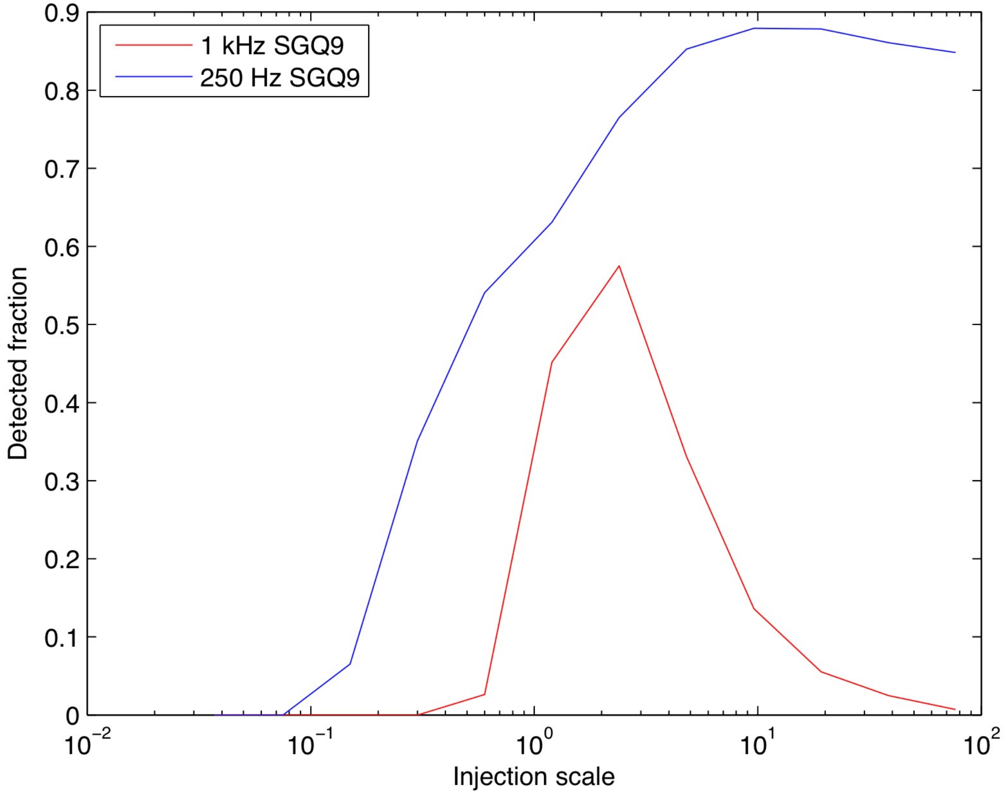
- Testing: expand the slow MC reference implementation/test codes to cover more of the functionality

High frequency signals

Detection statistic seems to work at low frequencies, but not at high frequencies

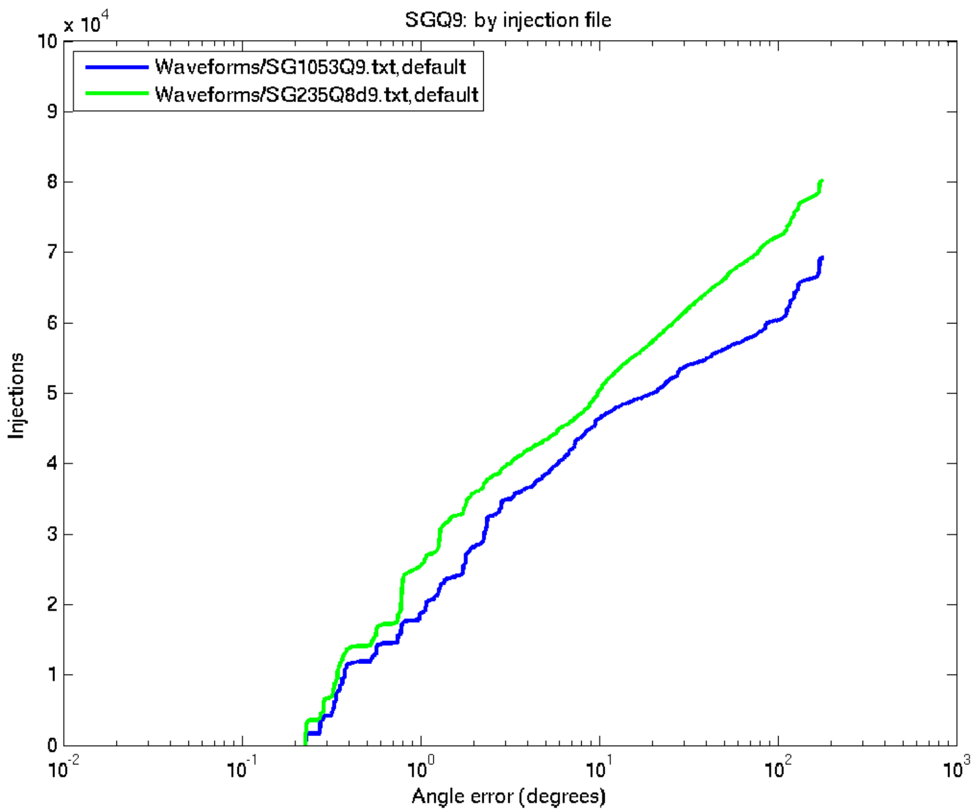
- Position reconstruction also slightly worsens with frequency, opposite to the expected scaling
- Testing:
  - Is integral sufficiently sampled? [some issues found but provisional "yes"]
  - Is inter-sample interpolation accurate enough?

High frequency signals



Fraction of injections more signal-like (according to probSignal) than best background event

Angle error for low and high frequency sine-Gaussians



White noise bursts

Relatively poor white noise burst performance

- Clustering may be modfied to preferentially decompose a patch with multiple degrees of freedom into long duration narrow tiles without frequency overlap, so that Bayesian signal model of one event in matched filter time series is true.

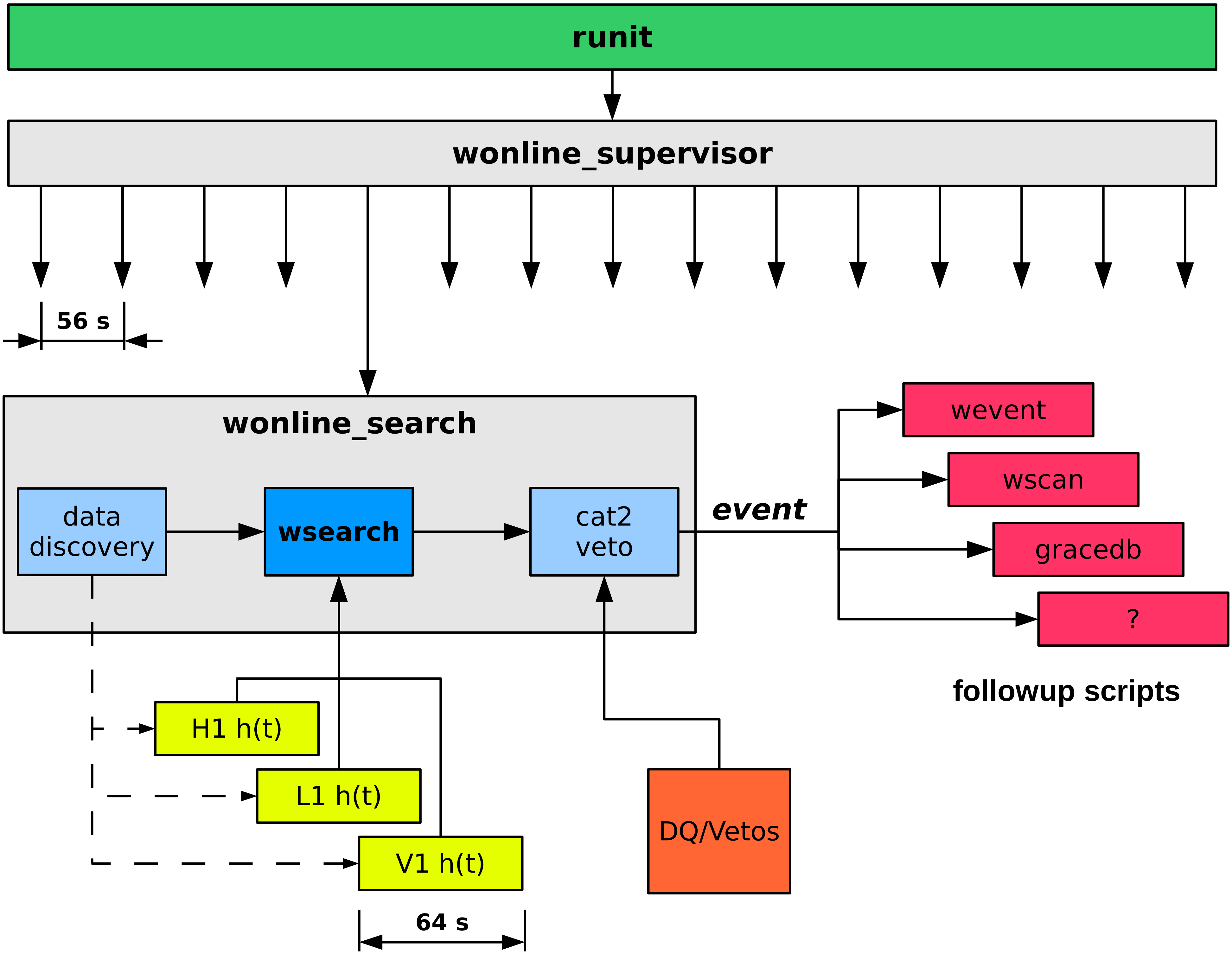
Conclusion

- No more new code, implementation complete
- Many features of the code work
  - Position reconstruction
  - Detection statistic at low-frequencies
- Some outstanding bugs, but plans to combat them:
  - Skymap calibration
    - Expand the slow MC reference implementation/test codes to cover more of the functionality
  - High frequency performance
    - Complete testing that the numerical integral's sampling is sufficient
    - Validate the inter-sample interpolation performance at high frequencies and SNRs
  - Relatively poor white noise burst performance
    - Alter the clustering algorithm to favor covering extended time-frequency patches with long-duration narrow-bandwidth tiles

Omega Rapid Online Analysis

Omega Rapid Online Analysis

- Full Bayesian coherent multi-detector analysis
- Analyses overlapping 64-second blocks, 4 second overlap
- Analyses new *online hoft* frames
- Analyses all in-lock data (*Cat1* from in frame state-vector)
- ~2.5 minute latency (depending on data availability)



runit
wonline_supervisor
56 s
wonline_search
data discovery
wsearch
cat2 veto
event
wevent
wscan
gracedb
?
followup scripts
H1 h(t)
L1 h(t)
V1 h(t)
DQ/Vetos
64 s

Status

- Running stably since beginning of S6
- Running on dedicated nodes at Caltech and AEI
- Integrated significance check against zero-lag and low-latency background
- Integrated *Cat2* veto check for event follow-ups
- Working Ω-event, Ω-scan and GraceDB notification followups
- Omega Online event web pages

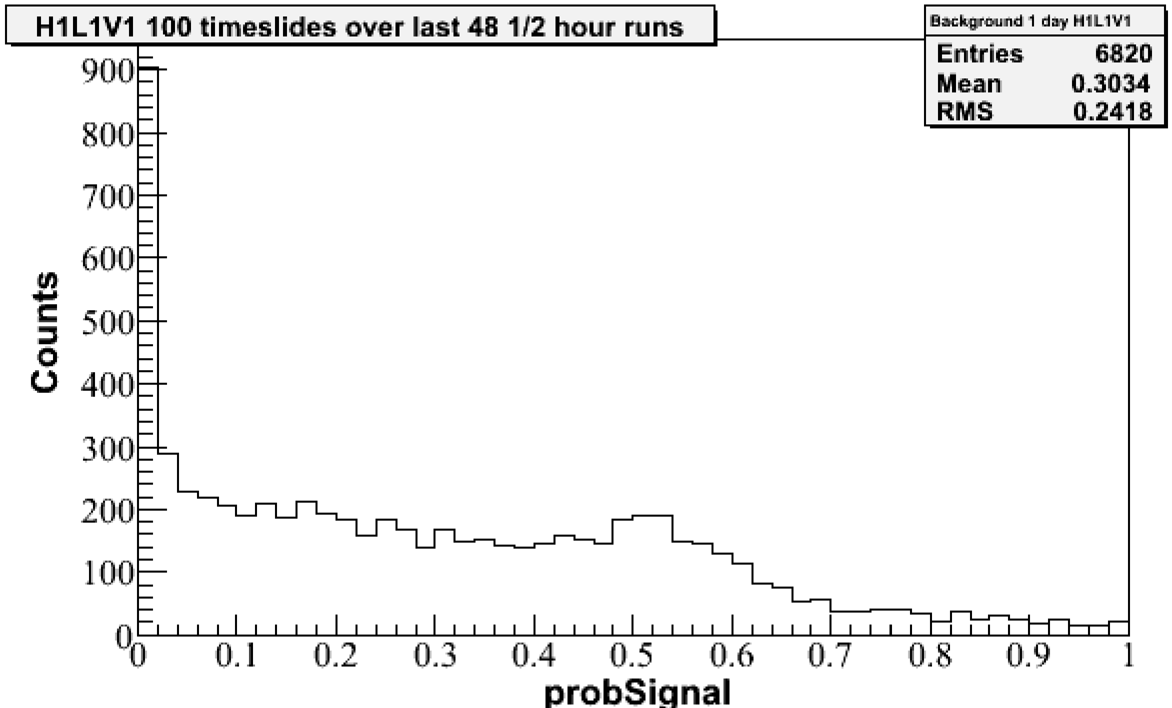
Significance checks

Two types of significance checks:

- Zero-lags:
  - Accumulation of all zero-lag events since beginning of run
- Background time lags:
  - 100 background time-lags every ½ hour
  - Check against most recent run

Background

- 100 background timelags are run at CIT in 30 minute chunks.
- Use identical parameters as online analysis
- Scripts collect results, apply online data quality, and make event lists available for rapid background assessment



- Runs with problems can be easily re-run manually
- Near real-time processing (when cluster cooperating)
- Omega Online background web pages

Background

- Tested successfully during S6a and tests are ongoing
- Still unclear how cluster usage will affect generation of background for the rest of S6

Future:

- Extra lags will be run for exceptional events (potential Swift followup candidates) to get more precise estimate of local background

New Omega projects

New Omega projects

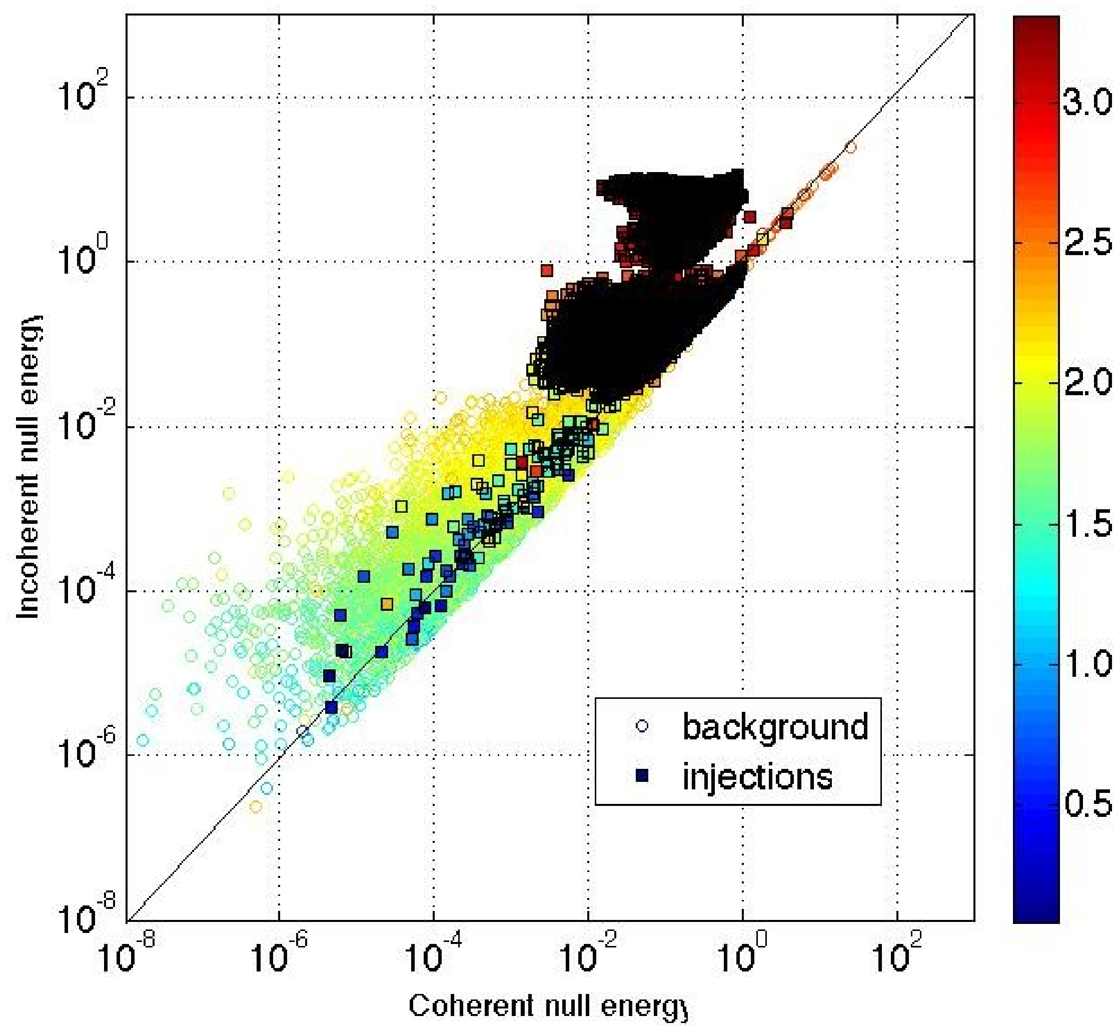
Various groups are working on interesting and promissing extensions to Omega pipeline:

- xCoherentCheck
- Maximun Entropy (MaxEnt)
- Boosted Decision Tree (BDT)
- Inspiral, Merger, Ringdown (IMR)

xCoherentCheck

Mark Edwards, Patrick Sutton (Cardiff)

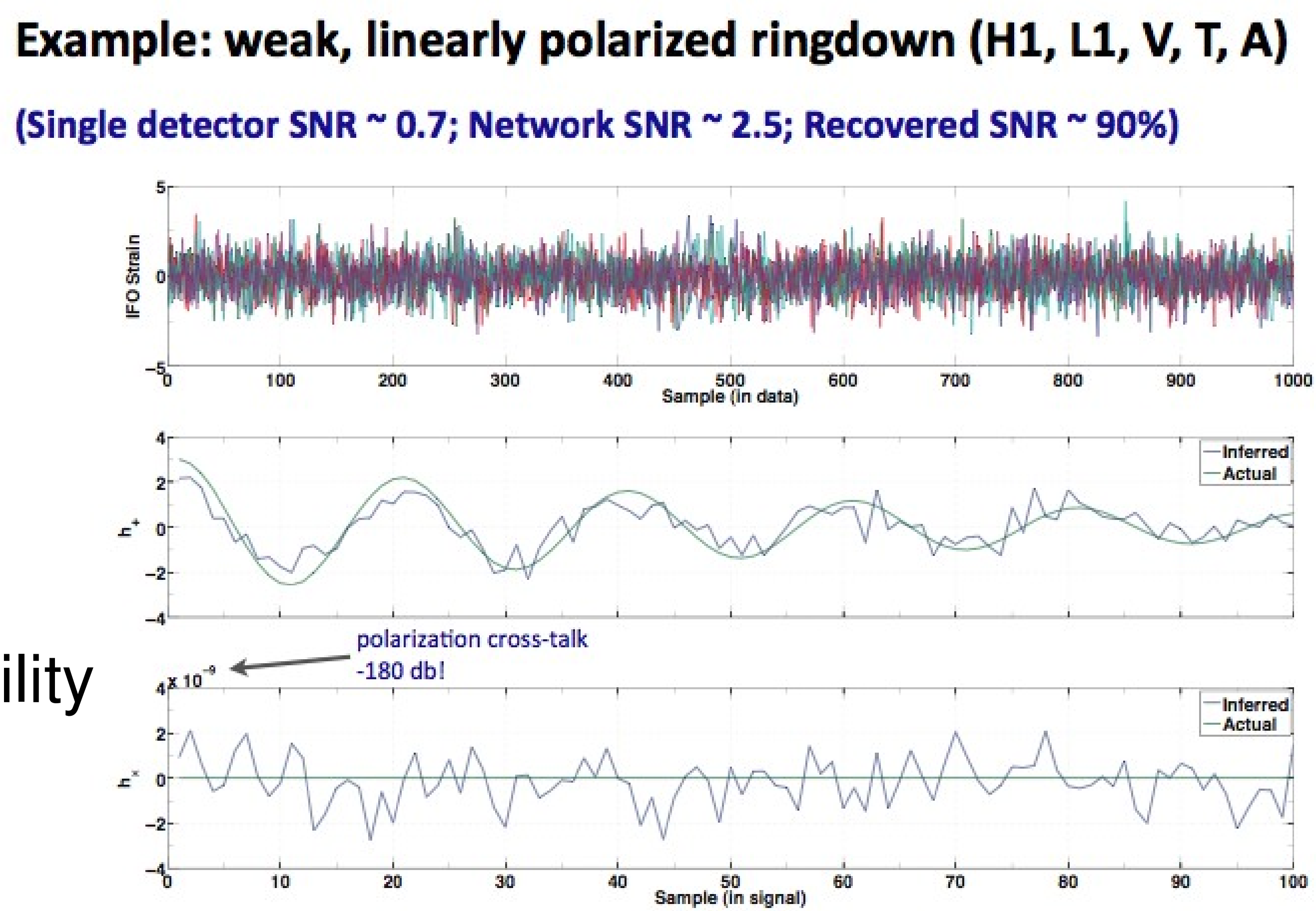
- Check event consistent by looking at coherent null stream (ala X-Pipeline)
- Currently investigating cuts: Injections (squares) generally lie above diagonal in *incoherent* vs. *coherent* null energy
- In integration (as event followup)



Maximum Entropy

Sam Finn, Meagan Lang, Ruxandra Bondarescu, Ravi Kopparapu and Ryan Fisher (Penn State), Tiffany Summerscales (Andrews U) and Andrea Lommen (Fraklin and Marshall College)

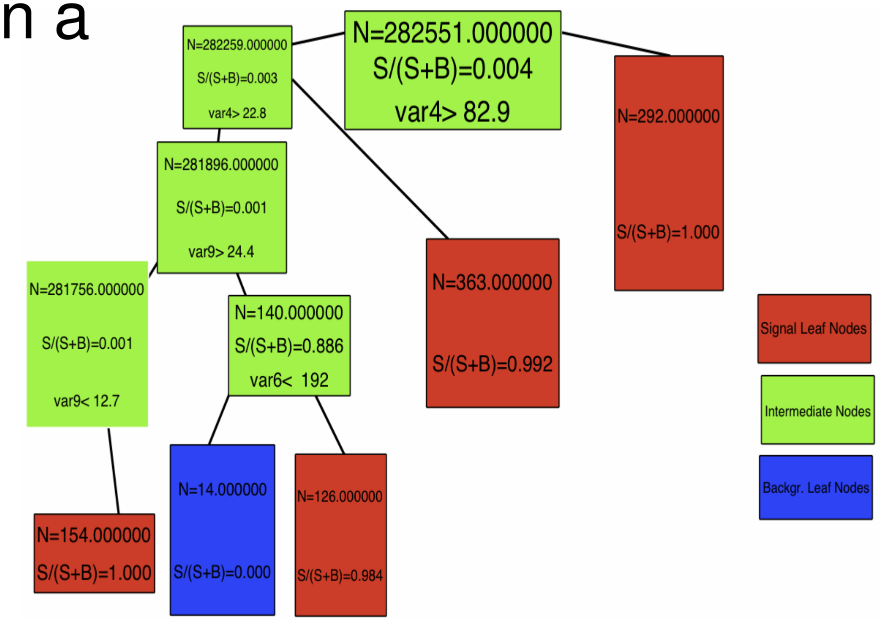
- Infer gravitational waveforms in noisy data from un-modeled sources via Bayesian deconvolution
- Finds probability density, P(h+,hx), and h for which this probability is maximized for given detector network
- In testing (as event followup)



Boosted Decision Tree

Satyanarayan Mohapatra, Laura Cadonati (UMass)

- Method for classifying events based on a series of “node-like” decisions(cuts) for a number of physical variables.
- BDT response is the combined vote of many individual decision trees, derived from the same training sample by boosting (re-weighting) events.



- BDT has been successfully tested to effectively suppresses the false alarm rate while preserving detection efficiency.
- BDT has robust performance when trained in one kind of waveform and tested in another.
- In testing (as event followup)

IMR

- Omega pipeline is being used in the IMR group as one of the Burst pipelines
- Run complete on last two months of S5 with EOBNR and PHENOM, IMR waveforms, in collaboration with Inspiral, Ringdown and CWB pipelines
- A common database infrastructure to compare the Omega pipeline triggers with other pipelines is being developed (see posters by Chad Hanna and talk by Chris Pankow)
- Comparison and coincidences of triggers among pipeline is being done with the above data sets