Searching for gravitational-wave bursts with the Q Pipeline

Shourov K. Chatterji

LIGO Science Seminar

2005 August 2

LIGO-G050332-00-Z

Searching for bursts

- Matched filter: project data stream onto known waveform

- Cross-correlation: project data stream from one detector onto data stream from another

- Time-frequency: project data stream onto a basis of waveforms designed to cover the targeted signal space

Searching for bursts

- For many potential burst sources, we do not have sufficiently accurate waveforms to permit matched filtering

- Cross-correlation requires at least two detectors and is complicated by the differing responses of non-aligned detectors

- Search for statistically significant excess signal energy in the time-frequency plane
- Can be extended to multiple detectors in a way that coherently accounts for differing response

Measurement basis

- Multiresolution basis of minimum uncertainty waveforms
- Resolves the most significant structure of arbitrary bursts
- Encompass maximal signal and minimal noise
- Provides the tightest possible time-frequency bounds, minimizing accidental coincidence between detectors
- Overcomplete basis allowed
- We are interested in detection, not reconstruction

Parameterization of unmodeled bursts

- Characteristic amplitude, $\|h\|$:

$$\|h\|^2 = \int_{-\infty}^{+\infty} |h(t)|^2 \, dt = \int_{-\infty}^{+\infty} |\tilde{h}(f)|^2 \, df$$

- Matched filter signal to noise ratio, ρ_0:

$$\rho_0^2 = \int_0^{\infty} \frac{4|\tilde{h}(f)|^2}{S_h(f)} \, df$$

- Normalized waveform, ψ

$$h(t) = \|h\| \psi(t)$$

$$\tilde{h}(f) = \|h\| \tilde{\psi}(f)$$

Parameterization of bursts

- Time, frequency, duration, and bandwidth:

$$t_c = \int_{-\infty}^{+\infty} t|\psi(t)|^2 \, dt \qquad \sigma_t^2 = \int_{-\infty}^{+\infty} (t - t_c)^2 |\psi(t)|^2 \, dt$$

$$f_c = 2\int_{0}^{+\infty} f|\tilde{\psi}(f)|^2 \, df \qquad \sigma_f^2 = 2\int_{0}^{+\infty} (f - f_c)^2 |\tilde{\psi}(f)|^2 \, df$$

- Time-frequency uncertainty: $\sigma_t \sigma_f \geq \frac{1}{4\pi}$

- Quality factor (aspect ratio): $Q = \frac{f_c}{\sigma_f}$

Tiling the search space

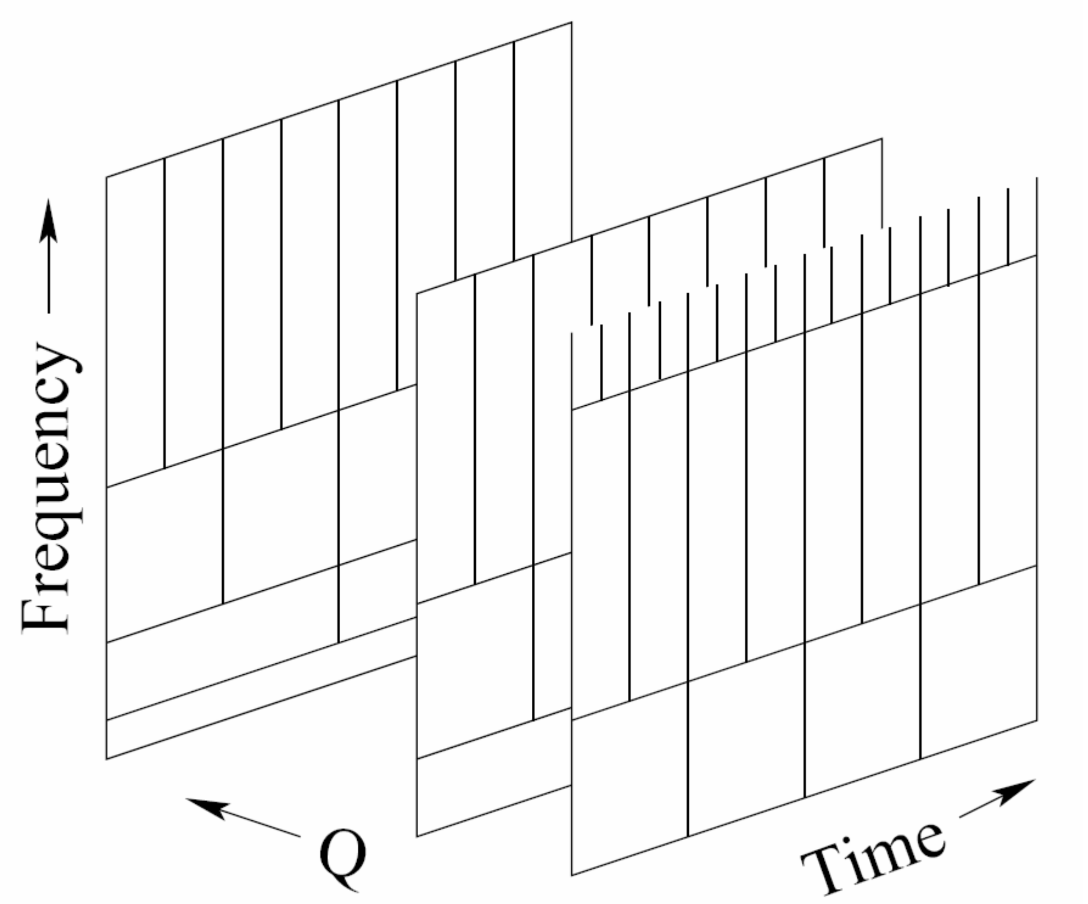
- Fractional signal energy lost due to mismatch between arbitrary minimum uncertainty burst and nearest basis function:

$$\mu(\delta\tau, \delta\phi, \delta Q) \simeq \frac{4\pi^2\phi^2}{Q^2}\,\delta\tau^2 + \frac{2+Q^2}{4\phi^2}\,\delta\phi^2 + \frac{1}{2Q^2}\,\delta Q^2 - \frac{1}{\phi Q}\delta\phi\,\delta Q$$

- Tile the targeted space of time, frequency, and Q with the minimum number of tiles to enforce a requested worst case energy loss
- Logarithmic spacing in Q
- Logarithmic spacing in frequency
- Linear spacing in time

Tiling the search space

- Naturally yields multiresolution basis
- Generalization of discrete wavelet tiling



Linear predictive whitening

- "Whiten" data by removing any predictable signal content
- Greatly simplified our subsequent statistical analysis
- Predict future values of time series

$$\tilde{x}[n] = \sum_{m=1}^{M} c[m]x[n-m]$$

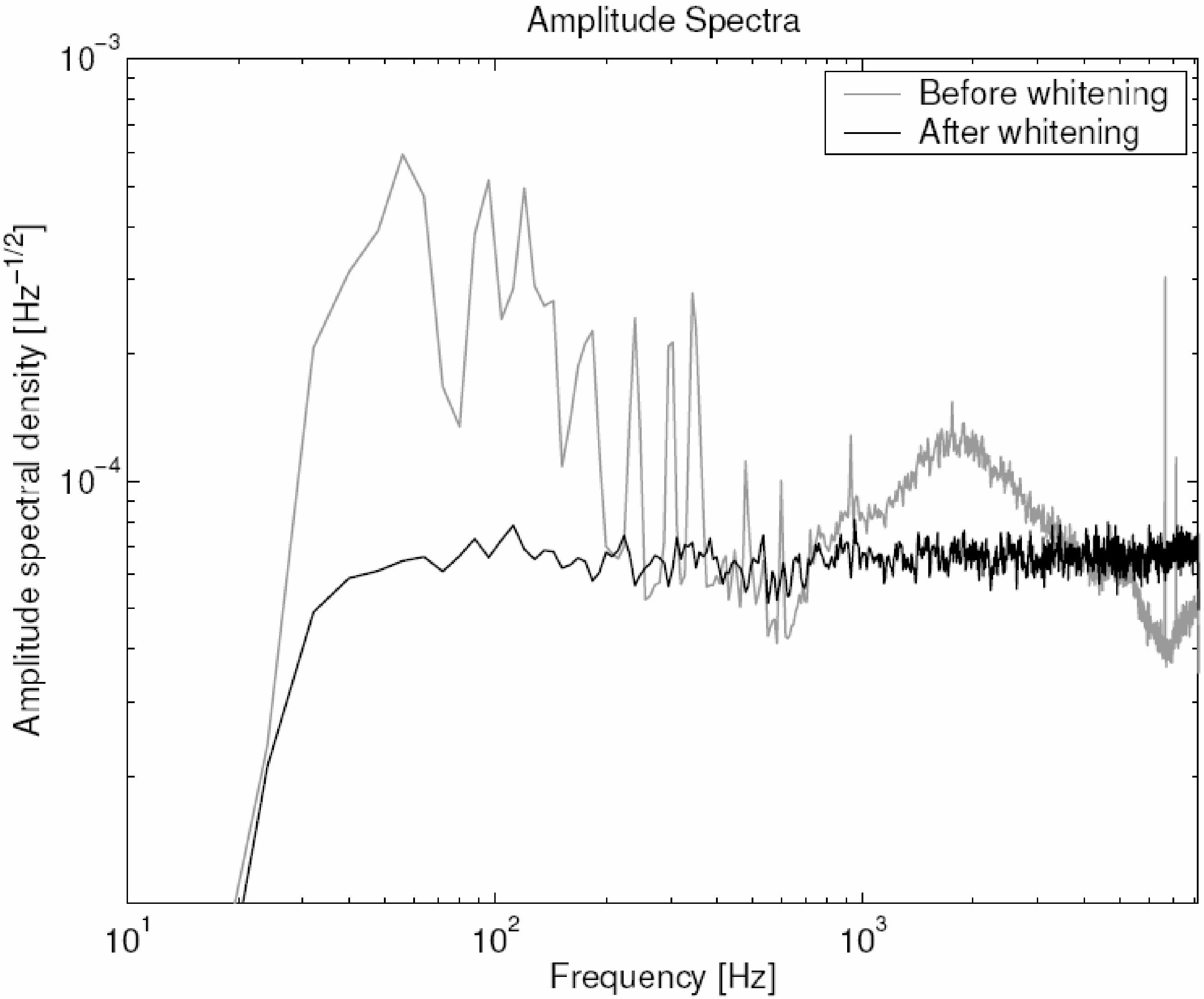
- Error signal is the signal of interest

$$e[n] = x[n] - \tilde{x}[n]$$

Linear predictive whitening

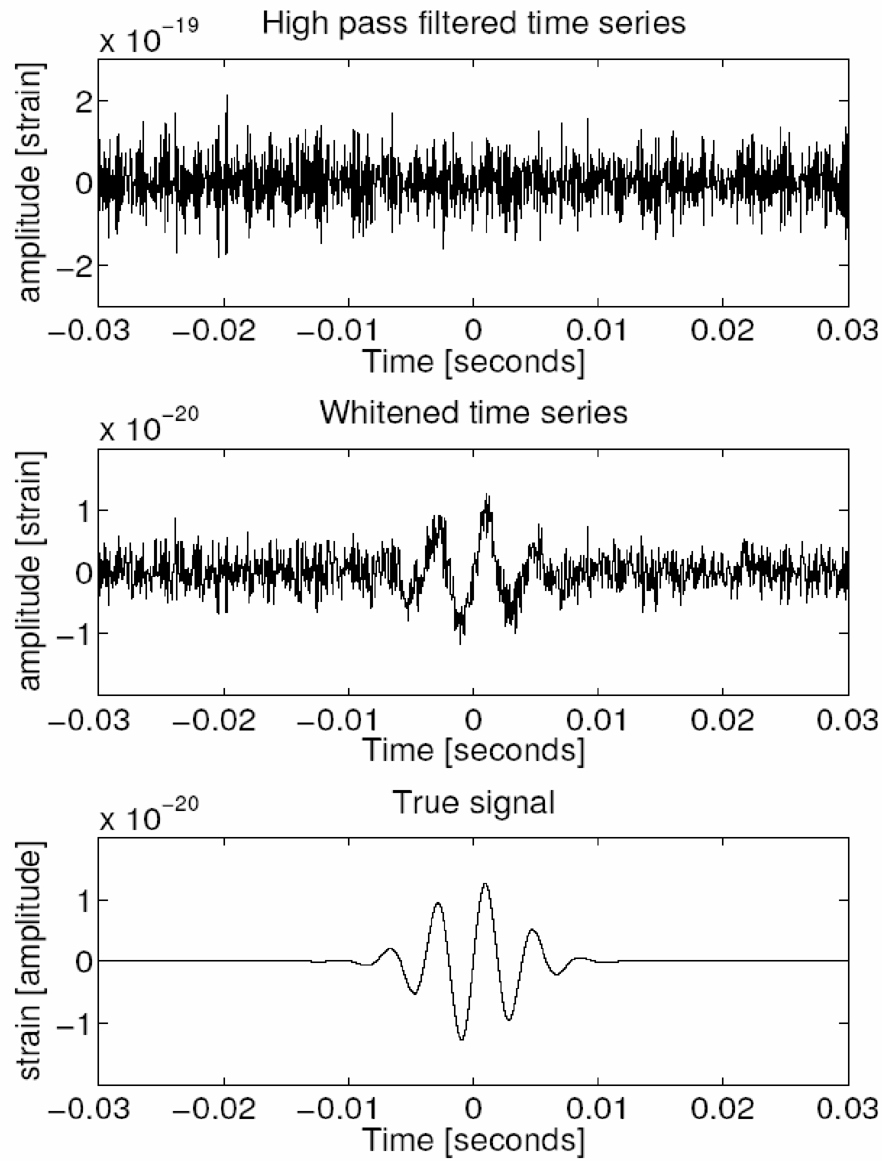
- Training: determine the filter coefficients by least squares minimization of the error signal
- Leads to the well known Yule-Walker equations
- The solution of this problem is well known, robust, and computationally efficient algorithms are available
- Filter order M should be longer than the longest basis function of the measurement
- Training time should be much longer than typical burst

Example whitened spectrum



Zero-phase whitening

- Linear predictive whitening introduces arbitrary phase delays between detectors that could destroy coincidence
- Zero-phase delay can be enforced by constructing a filter with symmetric coefficients that yields the same magnitude response
- Zero-phase high pass filtering is also possible by first causal and then acuasal filtering of the data



The Q transform

- Project whitened data onto multiresolution basis of minimum uncertainty waveforms

$$X(\tau, \phi, Q) = \int_{-\infty}^{+\infty} x(t)\, w(t-\tau, \phi, Q)\, e^{-i2\pi\phi t}\, dt$$

- Alternative frequency domain formalism allows for efficiency computation using the FFT

$$X(\tau, \phi, Q) = \int_{-\infty}^{+\infty} \tilde{x}(f+\phi)\, \tilde{w}^*(f, \phi, Q)\, e^{+i2\pi f\tau}\, df$$

- Frequency domain bi-square window

$$\tilde{w}(f, \phi, Q) = \begin{cases} \left(\frac{315}{128\sqrt{11}}\frac{Q}{\phi}\right)^{1/2} \left[1 - \left(\frac{fQ}{\phi\sqrt{11}}\right)^2\right]^2 & |f| < \frac{\phi\sqrt{11}}{Q} \\ 0 & \text{otherwise} \end{cases}$$

The Q transform

- Normalized to return characteristic amplitude of well localized bursts

$$\int_{-\infty}^{+\infty} h(t;\tau,\phi,Q)\, w(t-\tau,\phi,Q)\, e^{-i2\pi\phi t}\, dt = \|h\| e^{i\theta}$$

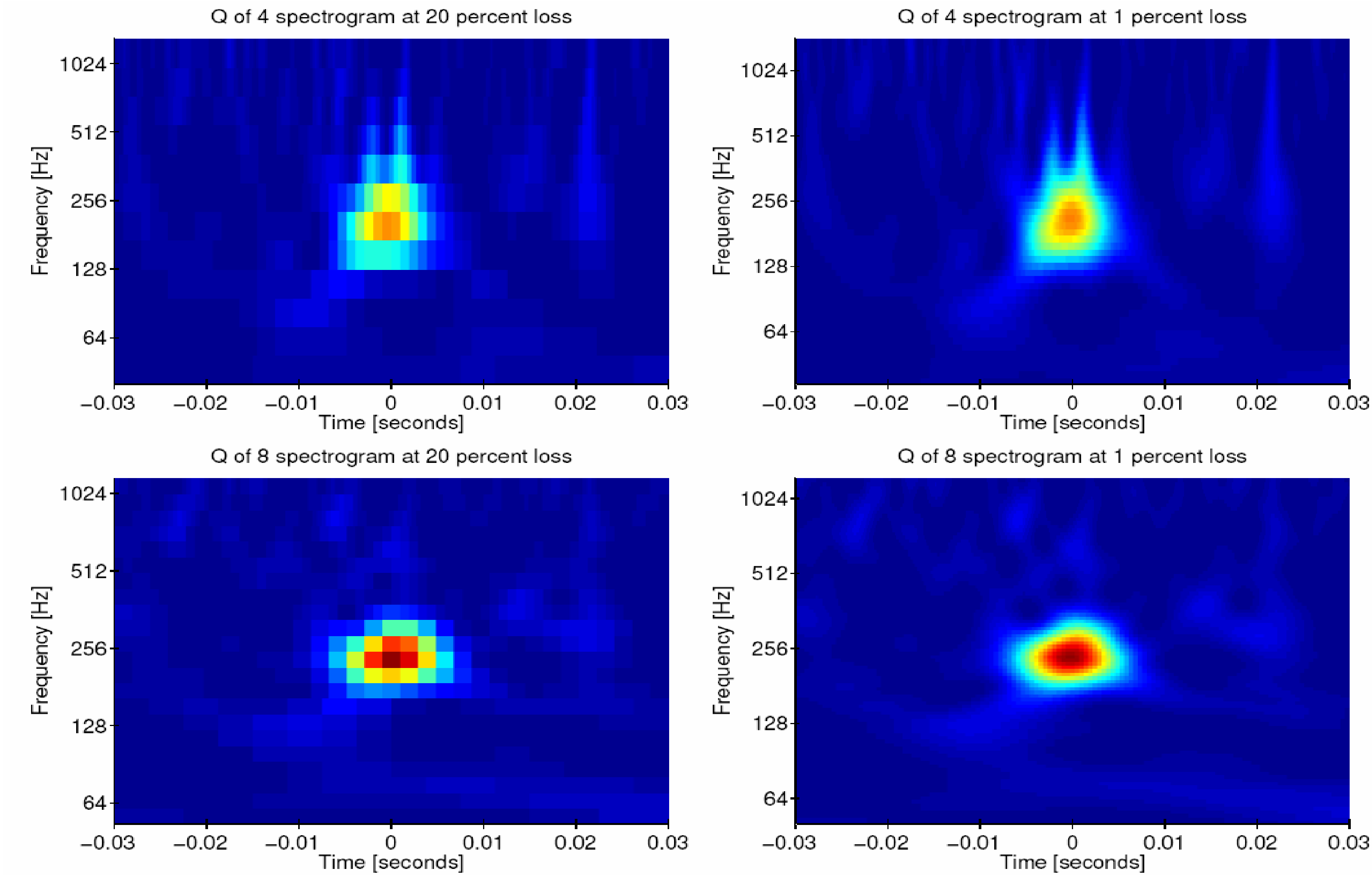
- Returns average power spectral density of detector noise if no signal is present

$$\left\langle |N(\tau,\phi,Q)|^2 \right\rangle = \frac{1}{2}\int_0^{\infty} S_n(f)\, |\tilde{w}(\phi-f)|^2\, df$$

- Alternative normalization permits recovery of the total energy of non-localized bursts

$$\int_0^{\infty}\int_{-\infty}^{+\infty} |X(\tau,\phi,Q)|^2\, d\tau\, d\phi = \|x\|^2$$

Example Q transform



White noise statistics

- Define the normalized energy,

$$Z = |X|^2 / \langle |X|^2 \rangle$$

- For white noise, this is exponentially distributed

$$f(Z)\,dZ = \exp(-Z)\,dZ$$

- Define the white noise significance:

$$P(Z' > Z) = \exp(-Z)$$

- Define the estimated signal to noise ratio:

$$\hat{\rho}^2 = Z - 1$$

Predicting performance

- Maximal false rate achieved if entire information content of data is tested:

$$\langle N(\hat{\rho}, f_s, T)\rangle = \exp(-\hat{\rho}^2 - 1) f_s T$$

- Measured signal to noise ratio:

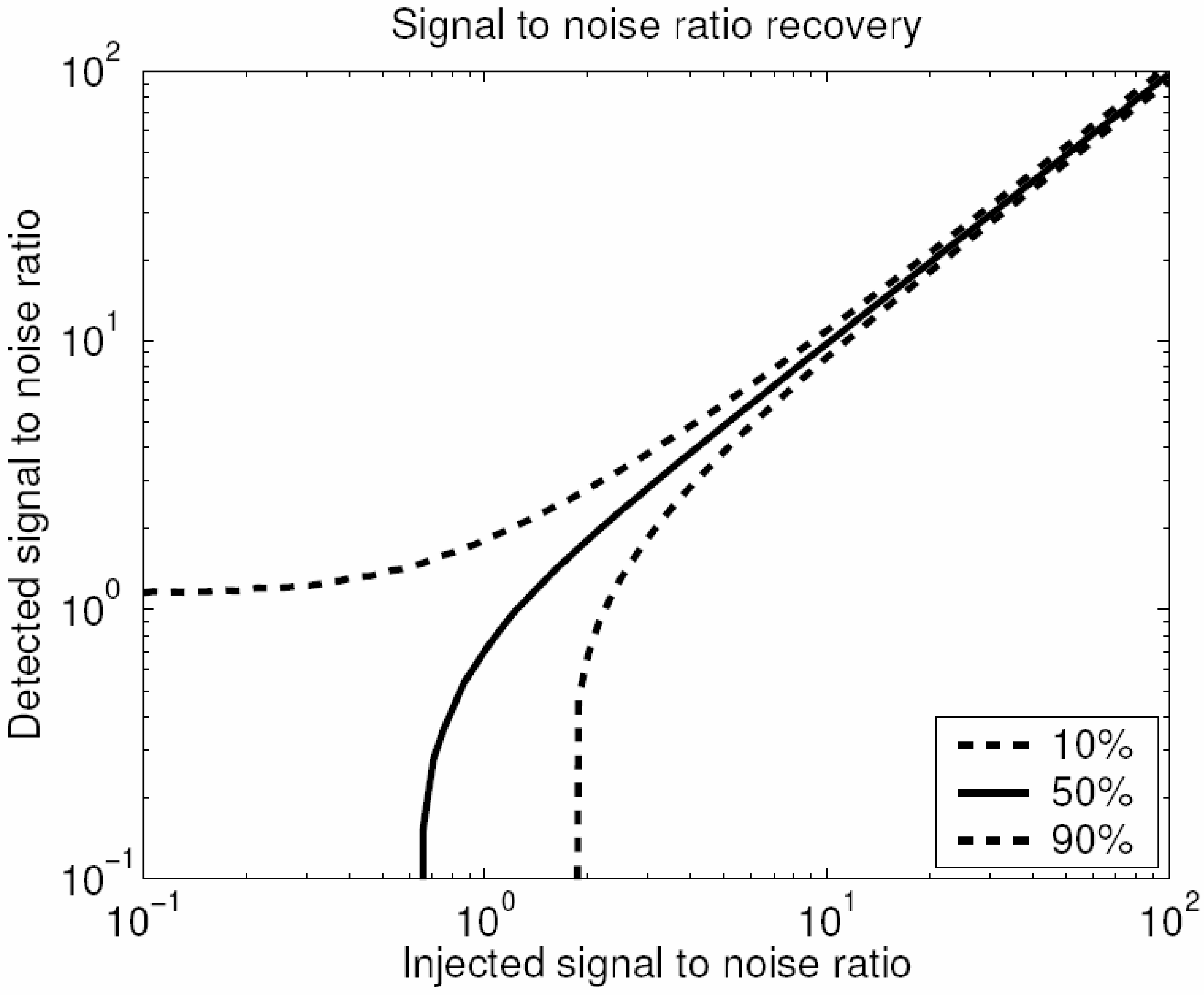
$$\hat{\rho}^2 = \frac{\|h\|^2 + |N(\tau,\phi,Q)|^2 + 2\|h\|\,|N(\tau,\phi,Q)|\cos\theta}{\langle |N(t,\phi,Q)|^2 \rangle} - 1$$

- True signal to noise ratio:

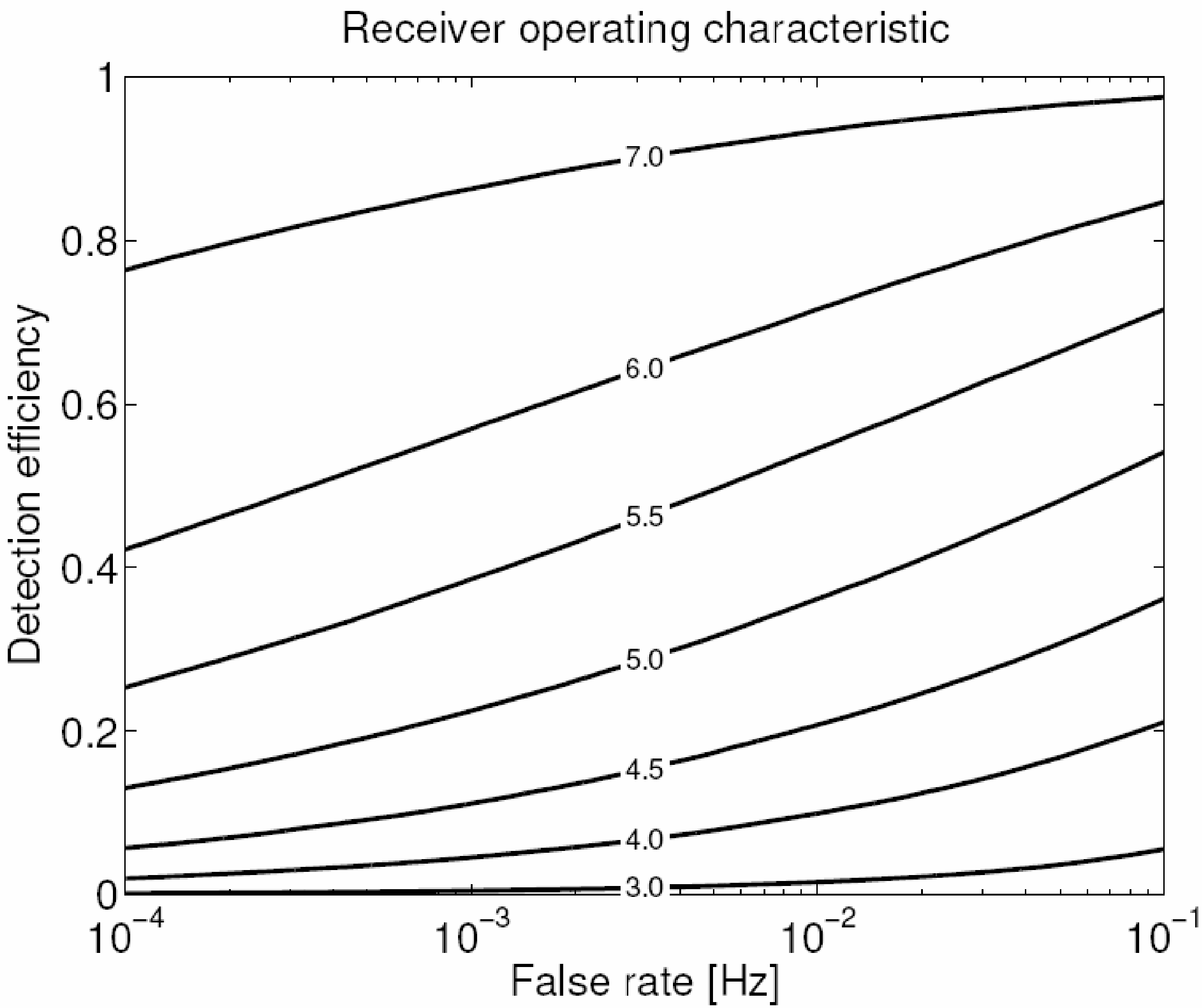
$$\rho^2 = \frac{\|h\|^2}{\langle |N(\tau,\phi,Q)|^2 \rangle}$$

- Monte Carlo expected performance based on exponential distribution of normalized energies and uniform distribution of relative phase

Ideal signal to noise ratio recovery



Ideal receiver operating characteristic



Coherent Q transform

- Gravitational wave signal in N collocated detectors:

$$X_n(\tau, \phi, Q) = h(\tau, \phi, Q) + N_n(\tau, \phi, Q)$$

- Form weighted linear combination of Q transforms

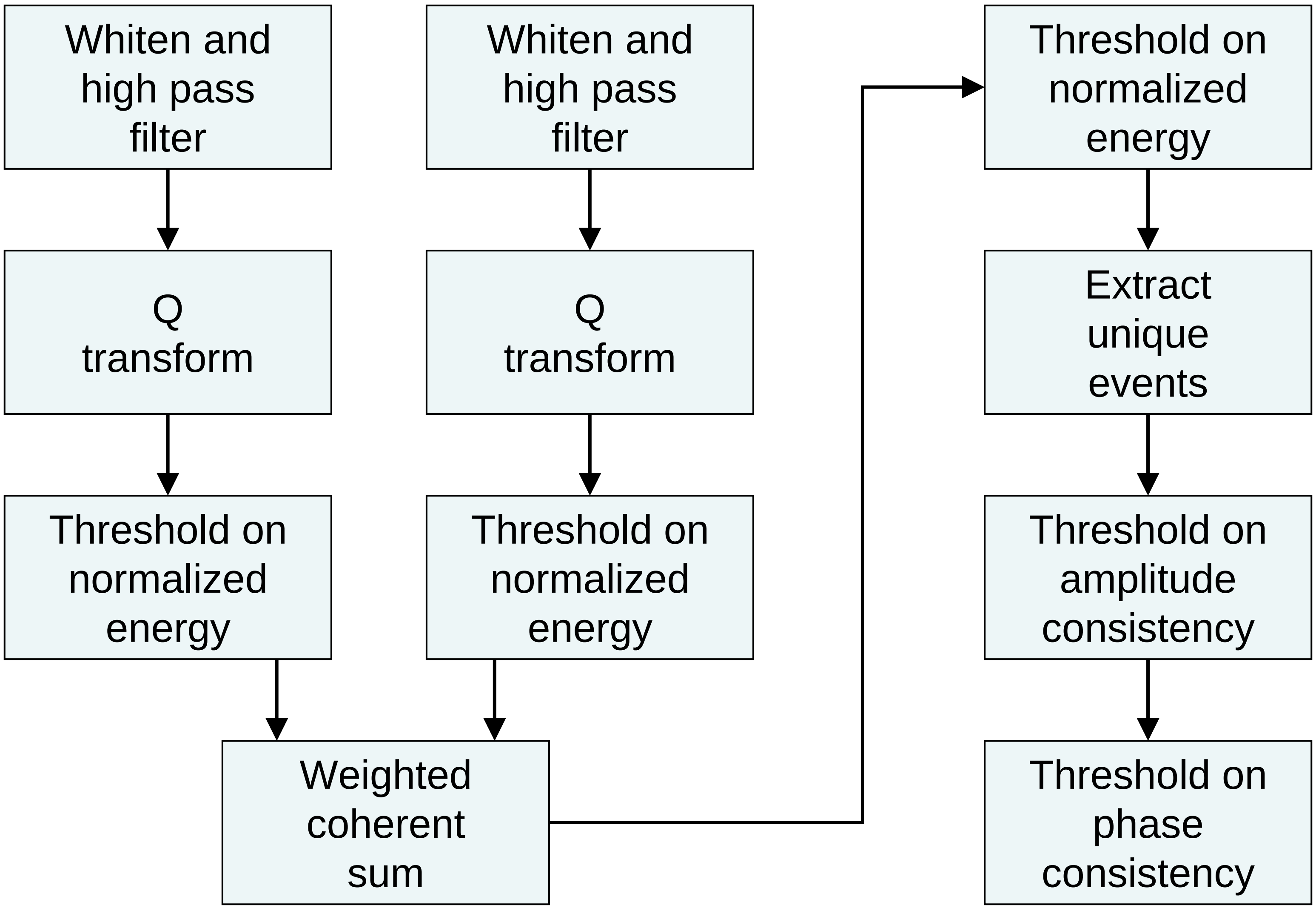
$$X^{(N)}(\tau, \phi, Q) = \sum_{n=1}^{N} C_n(\tau, \phi, Q) X_n(\tau, \phi, Q)$$

- Determine weighting coefficients to maximize expected signal to noise ratio:

$$C_n(\tau, \phi, Q) = \left(\sum_{m=1}^{N} \frac{1}{\langle |N_m(\tau, \phi, Q)|^2 \rangle} \right)^{-1} \frac{1}{\langle |N_n(\tau, \phi, Q)|^2 \rangle}$$

- Coherently combines Q transform from collocated detectors while taking into account frequency dependent differences in their sensitivity

Coherent Q analysis pipeline



Implementation

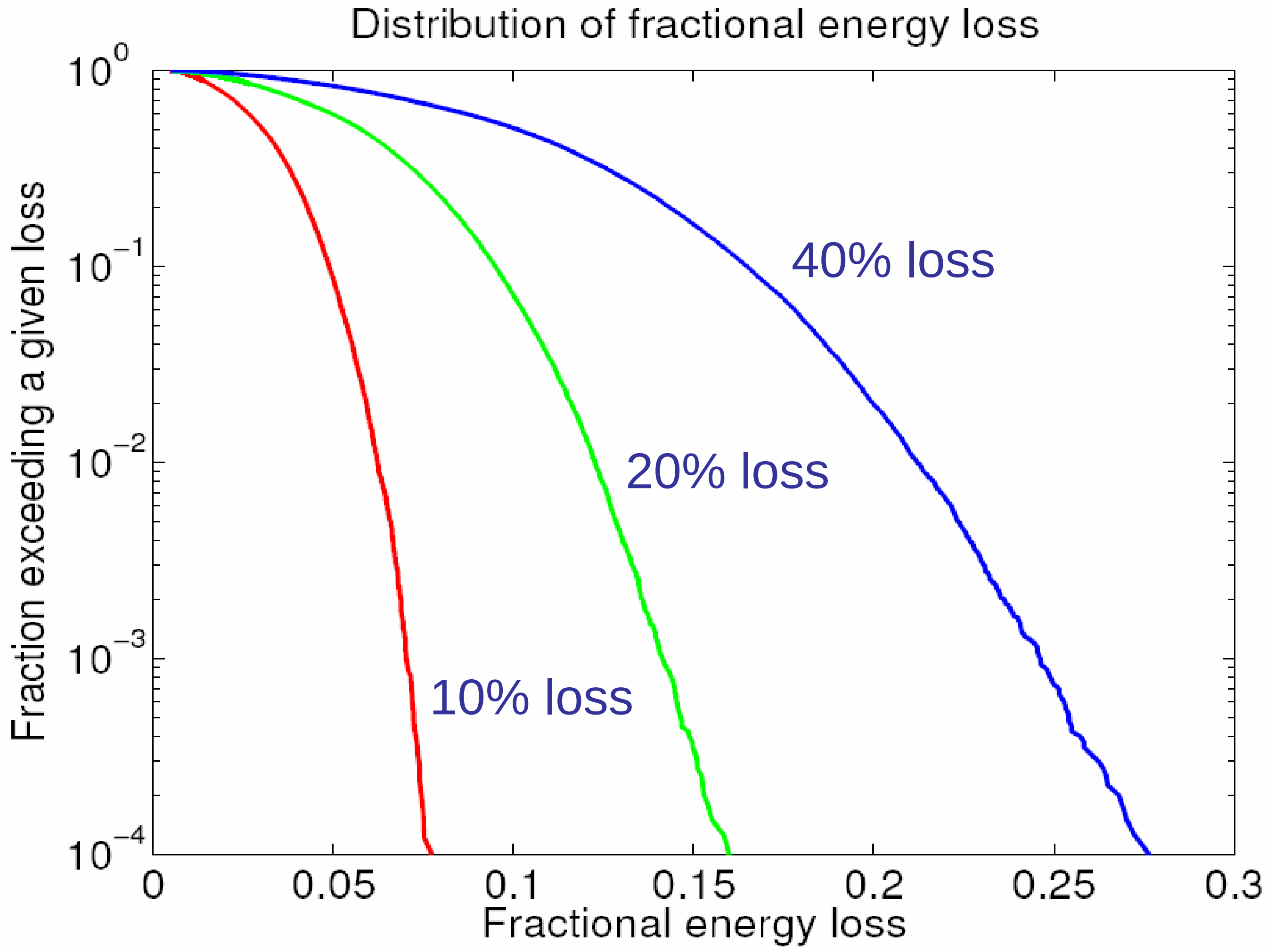
- Implemented in Matlab
- Compiled into stand alone executable
- Runs in 1.75 times faster than real time on a single 2.66 GHz Intel Xeon processor
- Foreground search performed in 1.5 hours on cluster of 290 dual processor machines using the Condor batch management system
- Code is freely available at http://ligo.mit.edu/~shourov/q/

Validation

- Does implementation perform as advertised?
- Simple tests of performance
- Compare with Monte Carlo predictions
- Inject sinusoidal Gaussian bursts with random center times, center frequencies, phases, Qs, and signal to noise ratios
- Into stationary white noise
- Into simulated detector noise

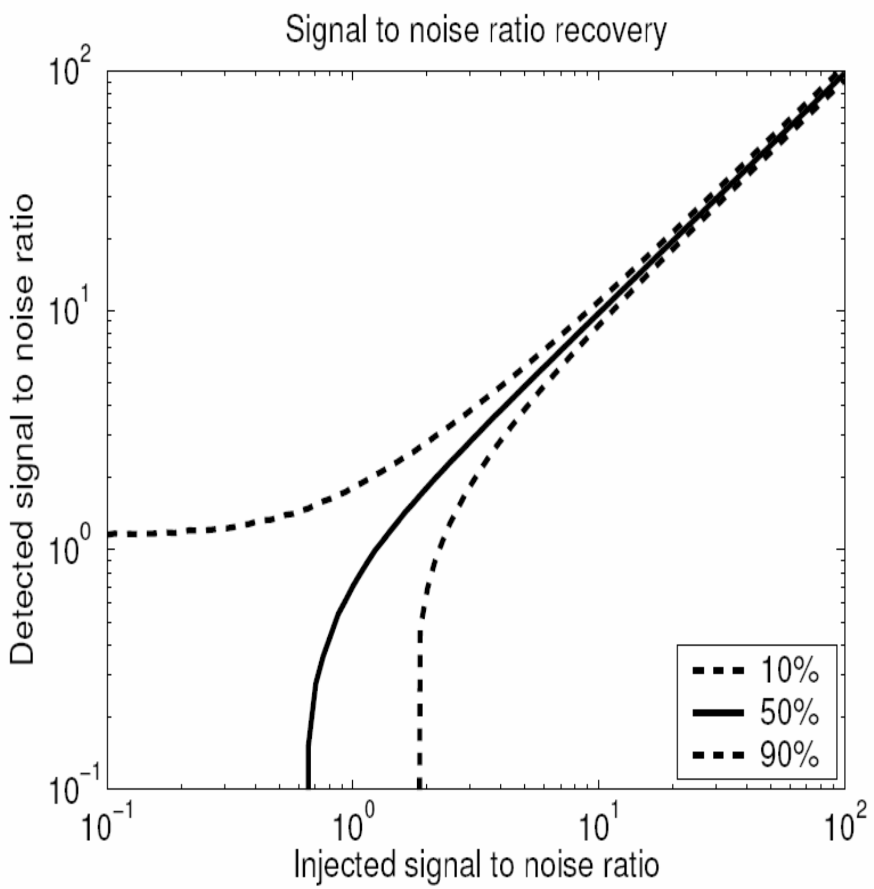
Tiling validation

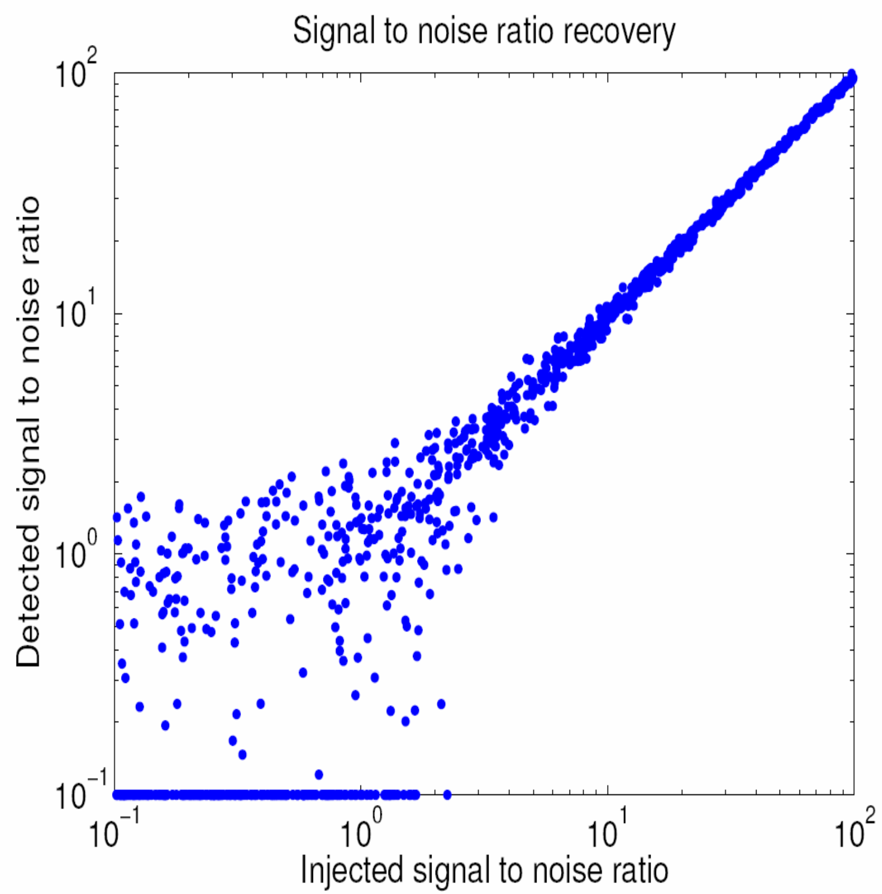
- Worst case energy loss is never exceeded in 40000 trials



Signal to noise ratio recovery

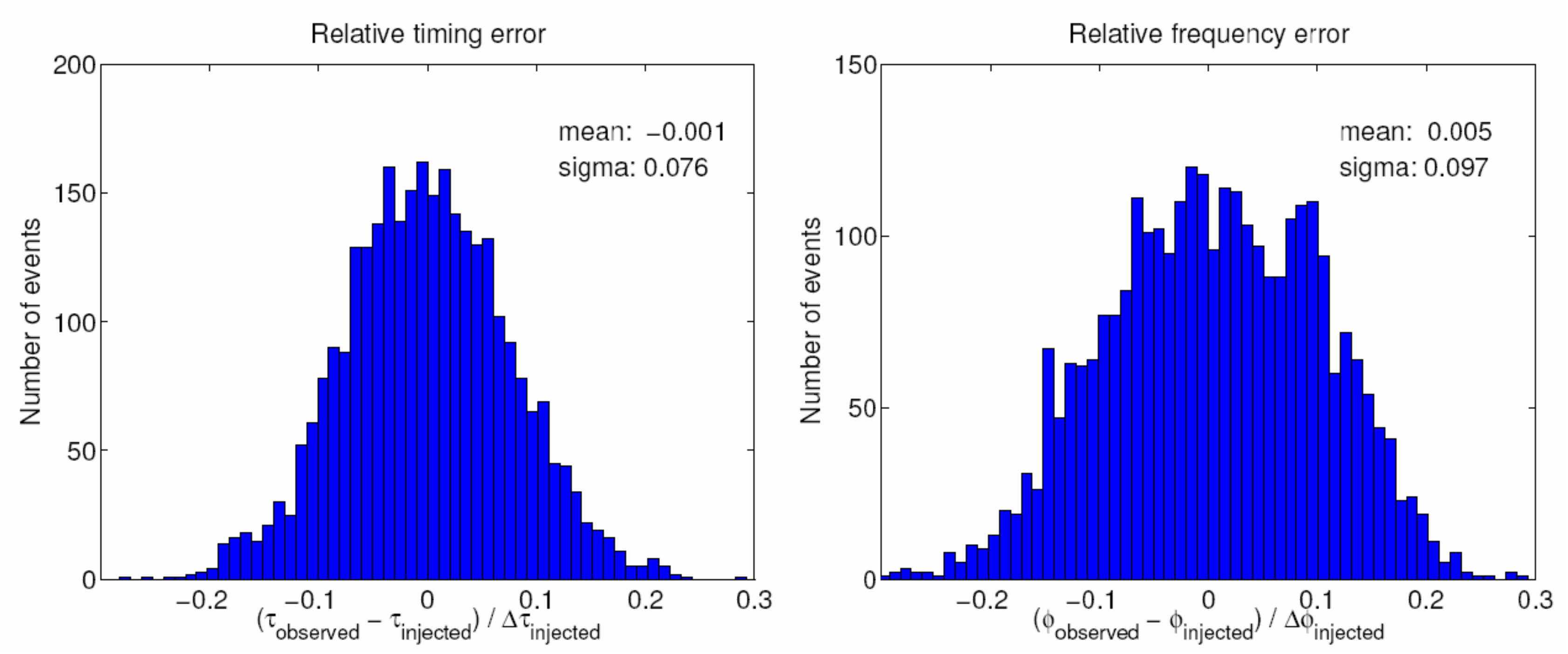
- Signal to noise recovery shows very good agreement with predicted performance





Measurement accuracy

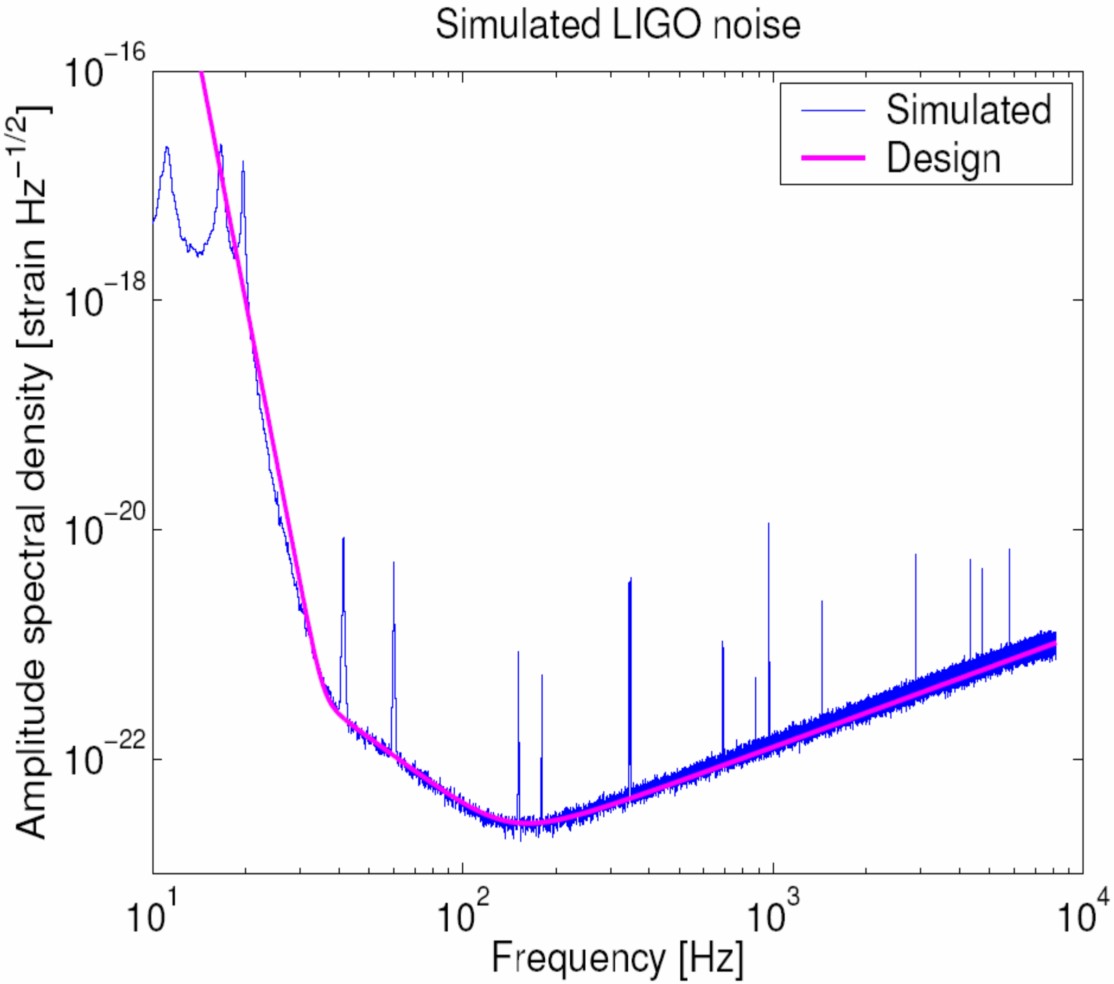
- Central time and frequency of sinusoidal Gaussians are recovered to within 10 percent of duration and bandwidth



- All signals injected with a signal to noise ratio of 10, but otherwise random parameters

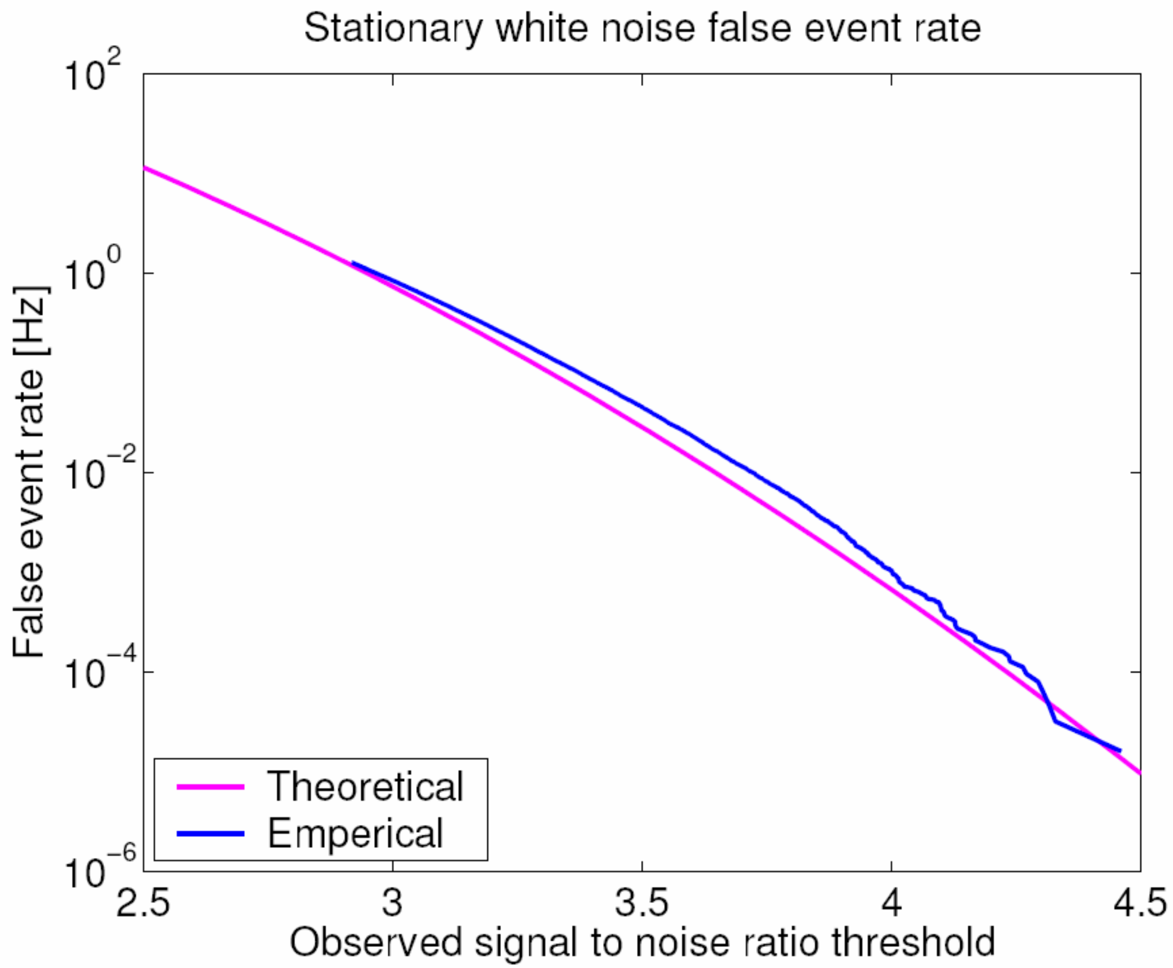
Simulated detector noise

- Simulated detector noise at LIGO design sensitivity
- Does not model non-stationary behavior of real detectors
- Provides end-to-end validation of pipeline, including linear predictive whitening
- Data set for benchmarking search algorithms



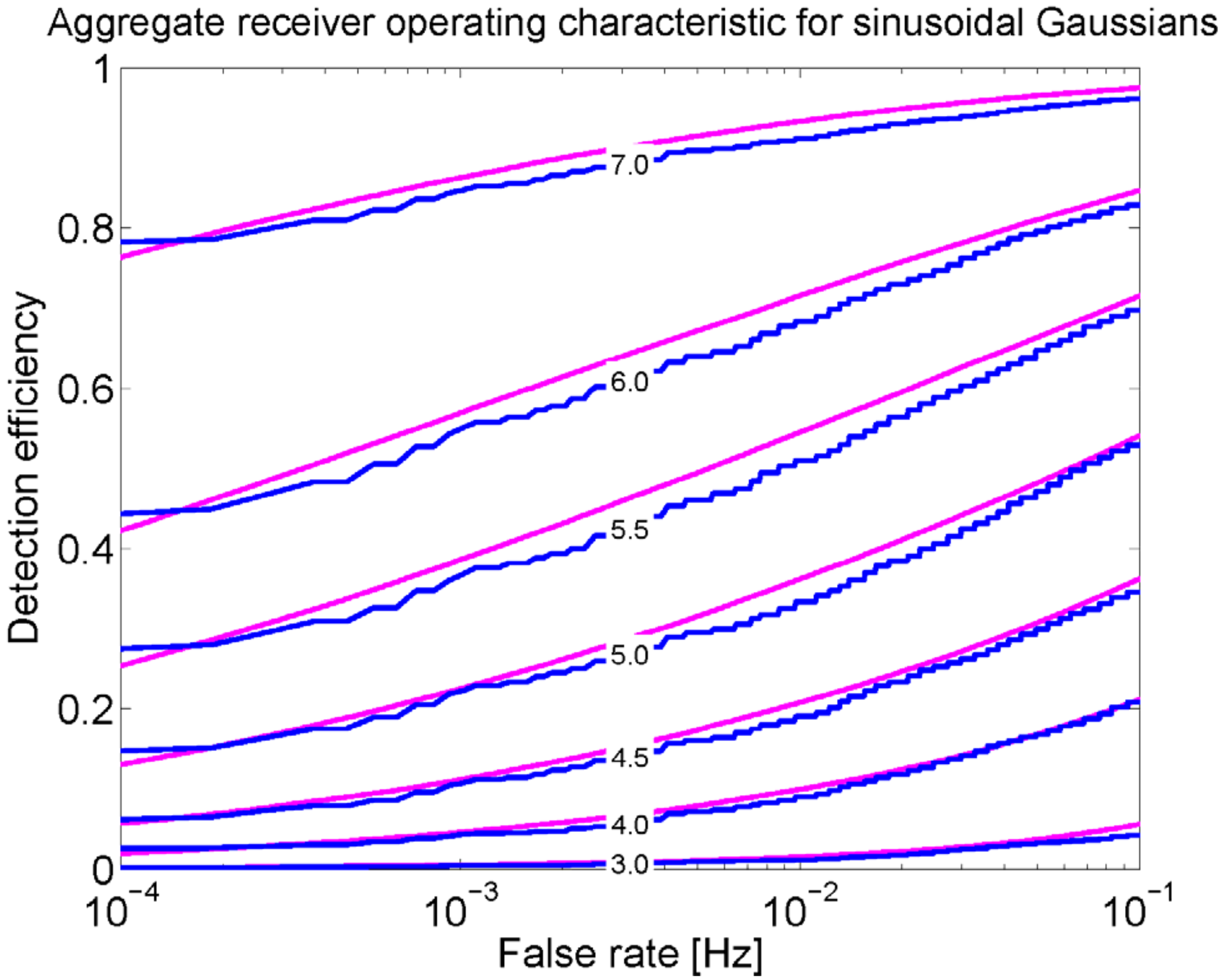
False detection rate

- Good agreement with maximal white noise false rate assuming full information content is tested



Receiver Operating Characteristic

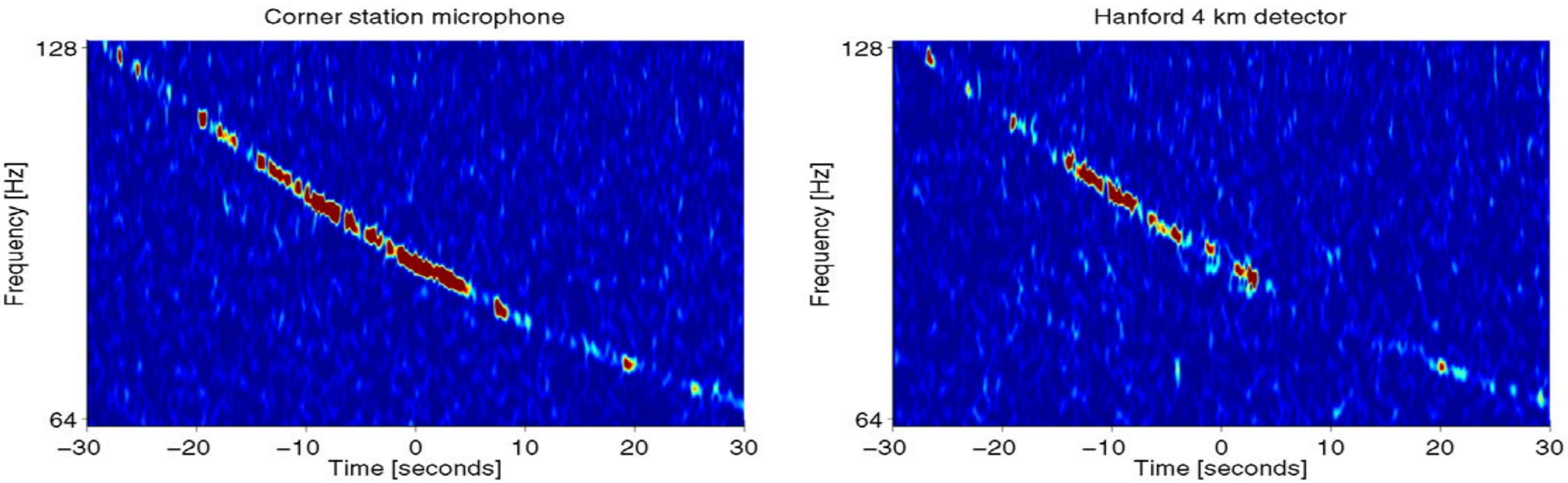
- Shows very good agreement with the performance of a templated matched filter search for sinusoidal Gaussians.



Example application

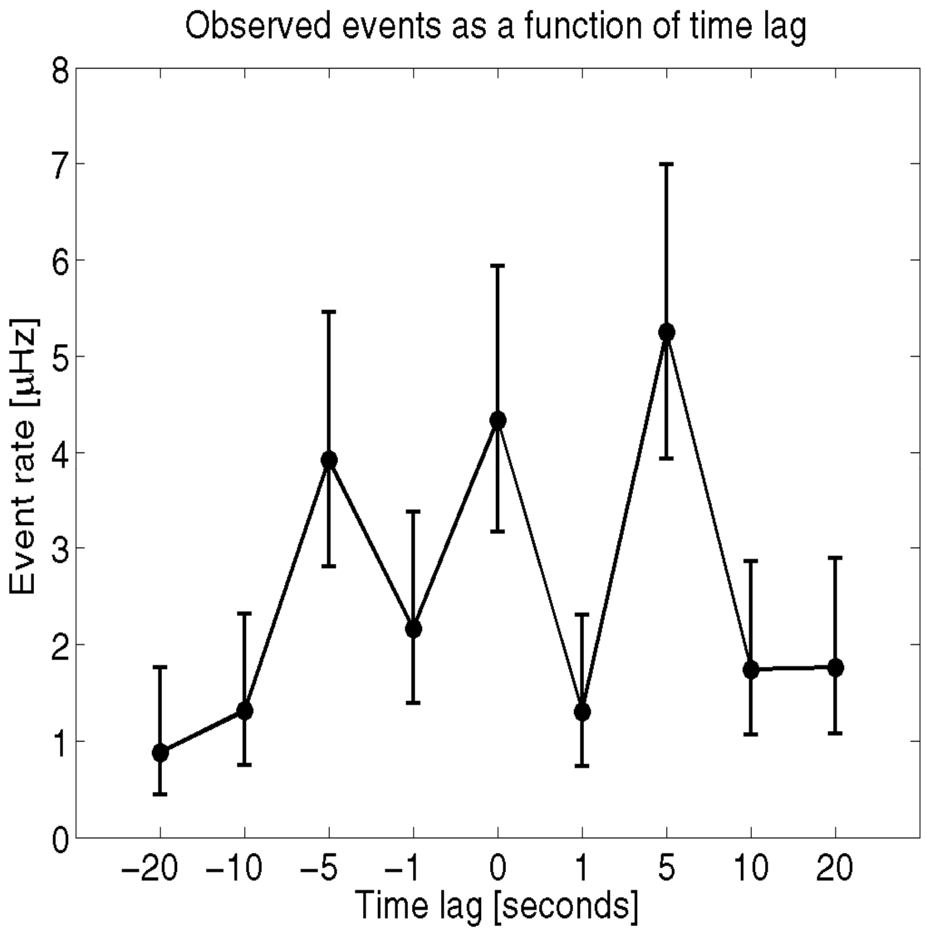
- Second LIGO science run
- Collocated double coincident Hanford data set
- Higher threshold required for false rate similar to triple coincident search
- Susceptible to correlated environmental noise
- Identical response permits coherent search and strict consistency tests
- ~ 2.3 times greater observation time than triple coincident search

Data quality and vetoes



- Acoustic coupling responsible for coincident events
- Periods of high acoustic noise excluded from analysis
- Q pipeline applied to microphone data to identify and exclude 290 additional acoustic events
- Also exclude times with missing calibration, anomalous detector noise, photodiode saturation, timing errors, etc.
- Remaining observation time is 645 hours

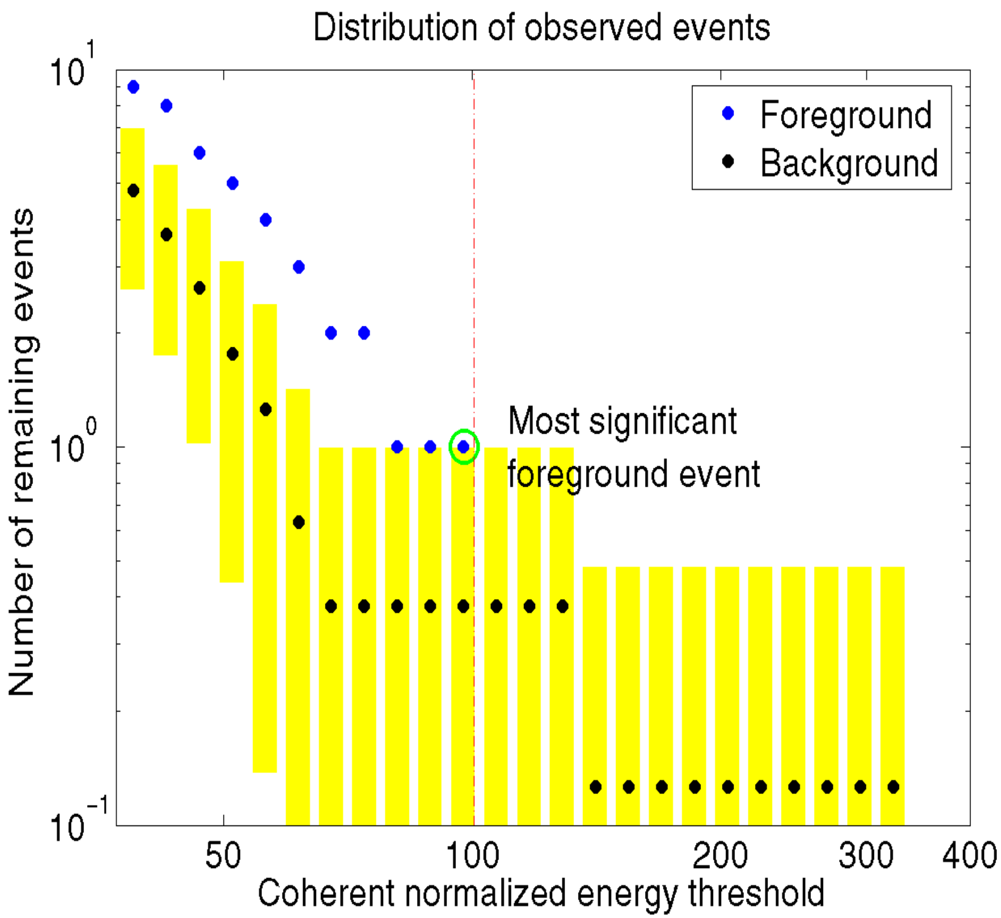
Background event rates



- Artificial time shifts used to estimate background event rate from random coincidence
- Does not estimate background event rate due to environment
- Statistical excess of events in unshifted foreground

- Interesting statistical excess of events at ±5 second lag
- Unknown environmental cause, possible microseism?

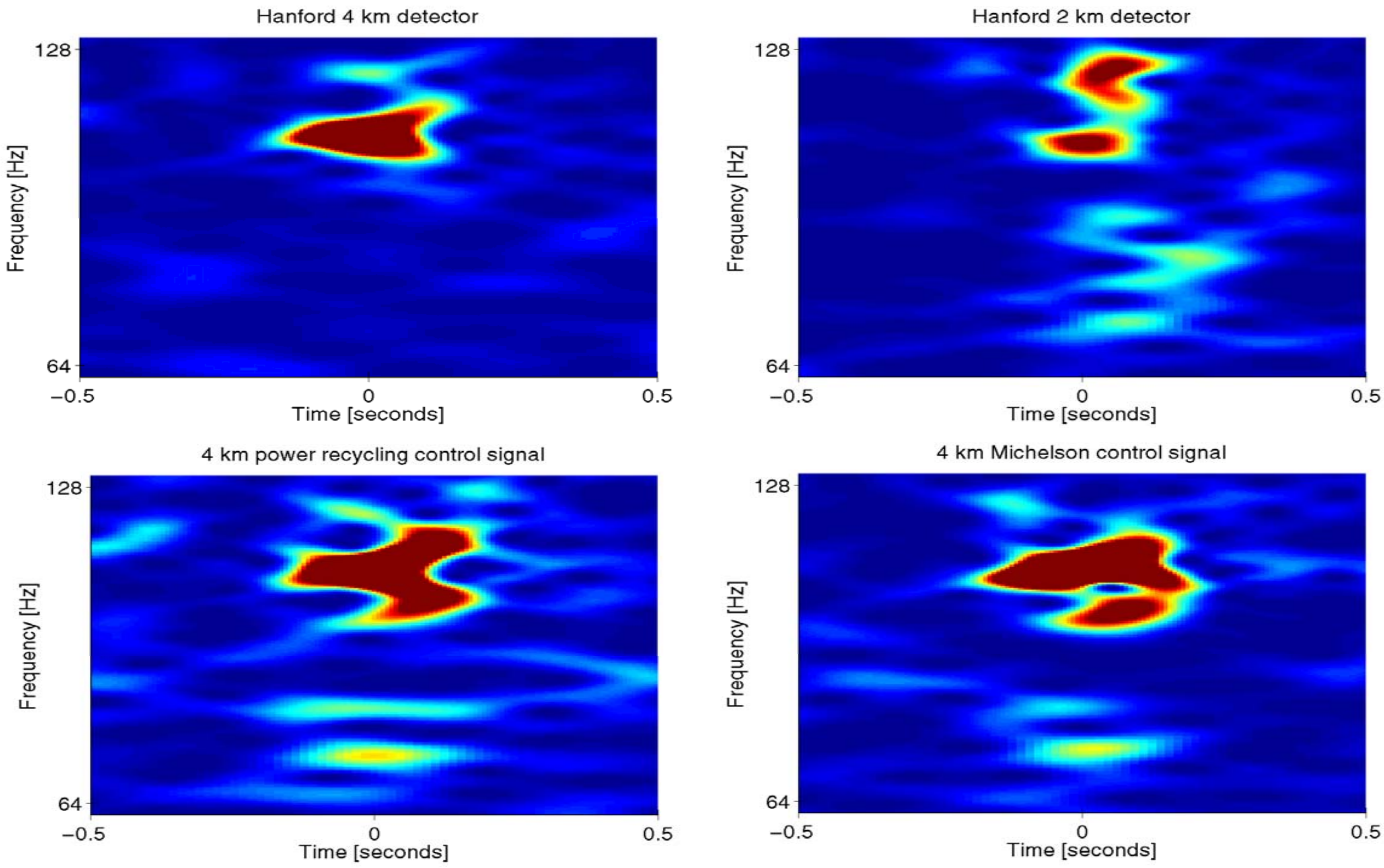
Foreground event rates



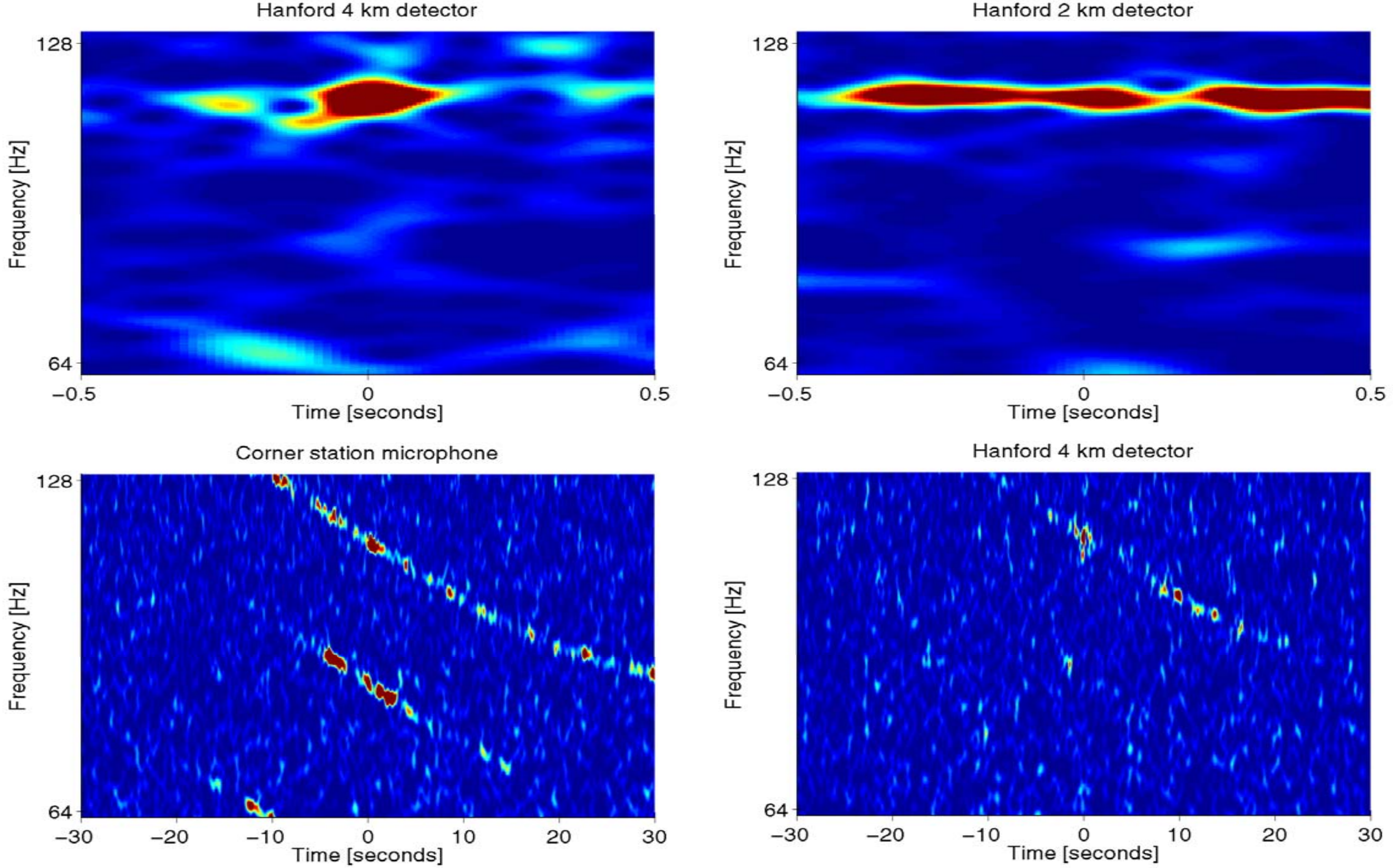
- 10 consistent events survive in the unshifted foreground
- Statistically significant excess foreground relative to accidental background
- Environmental origin, gravitational or otherwise

- The most significant event defines the search sensitivity
- But, first check to see if they are gravitational waves!

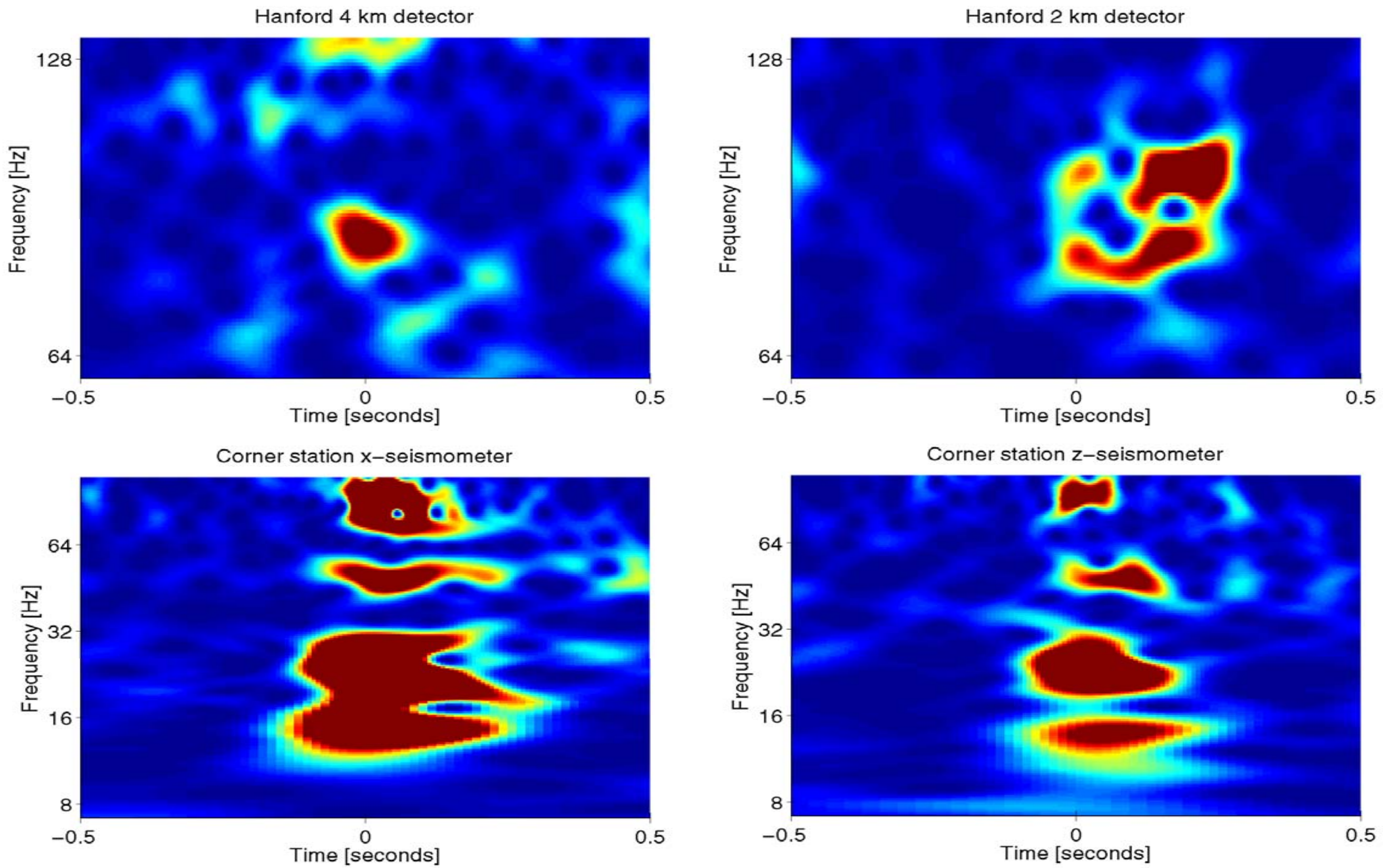
Most significant event: instrument artifact



Fourth most significant event: acoustic



Eighth most significant event: seismic



Loudest event statistic

- No gravitational-wave bursts are found!
- What was the sensitivity of the search?
- Determine frequentist upper bound on the rate of gravitational-wave bursts from an assumed population.
- Based on detection efficiency of population at the normalized energy threshold of the "loudest event"
- If we repeat the experiment, the stated upper bound exceeds the true rate in p percent of experiments:

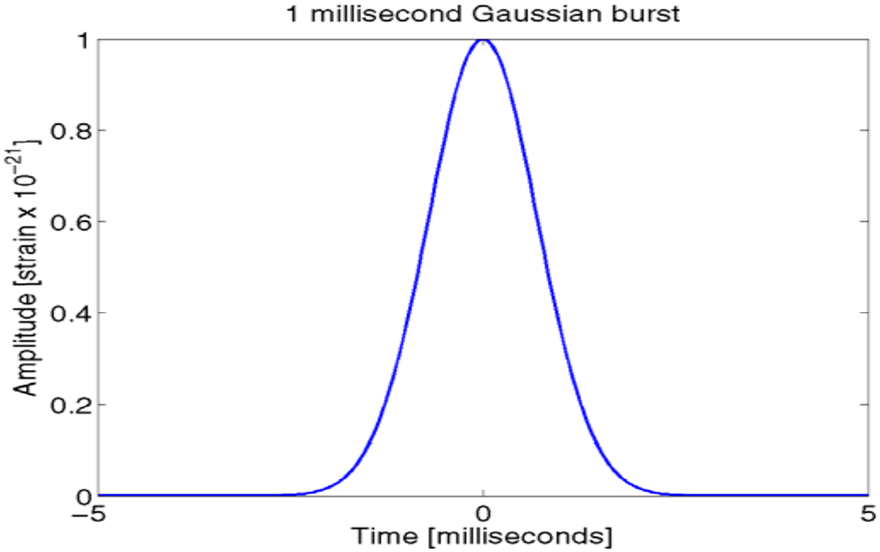
$$r_p = \frac{-\ln(1-p)}{T\varepsilon(Z)}$$

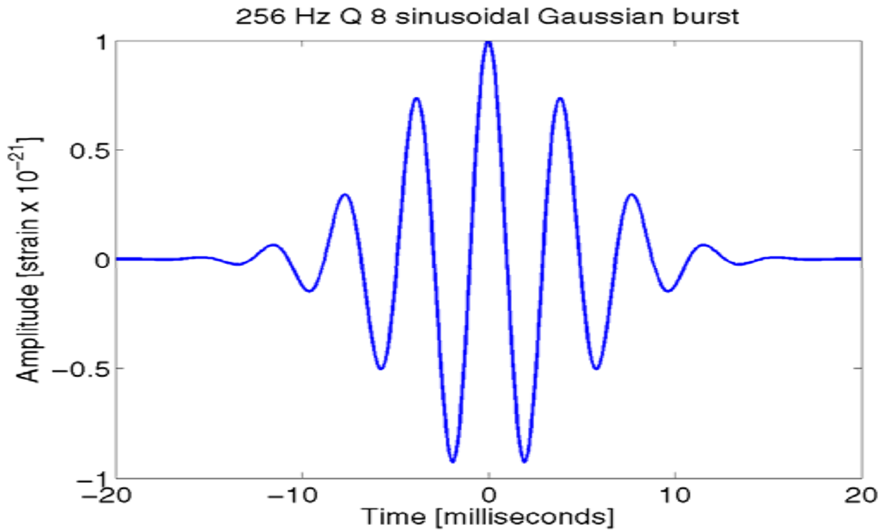
Interpreted upper limits

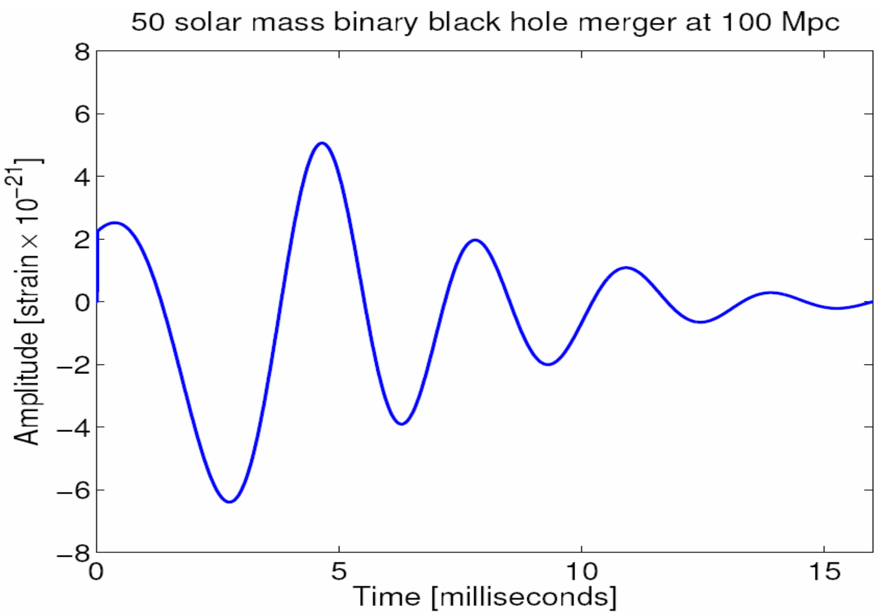
- Isotropic populations of identical bursts:

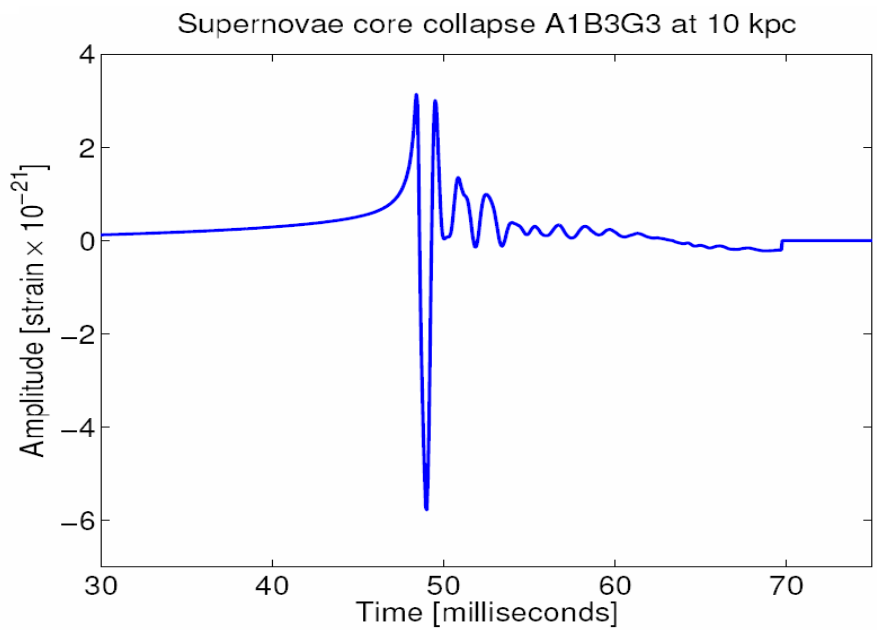
- Simple Gaussian bursts
- Sinusoidal Gaussian bursts
- Simulated black hole merger waveforms from Baker, et al.
- Simulated core collapse waveforms from Zwerger Mueller, et al., Dimmelmeir, et al., and Ott et al.

Simulated gravitational-wave bursts

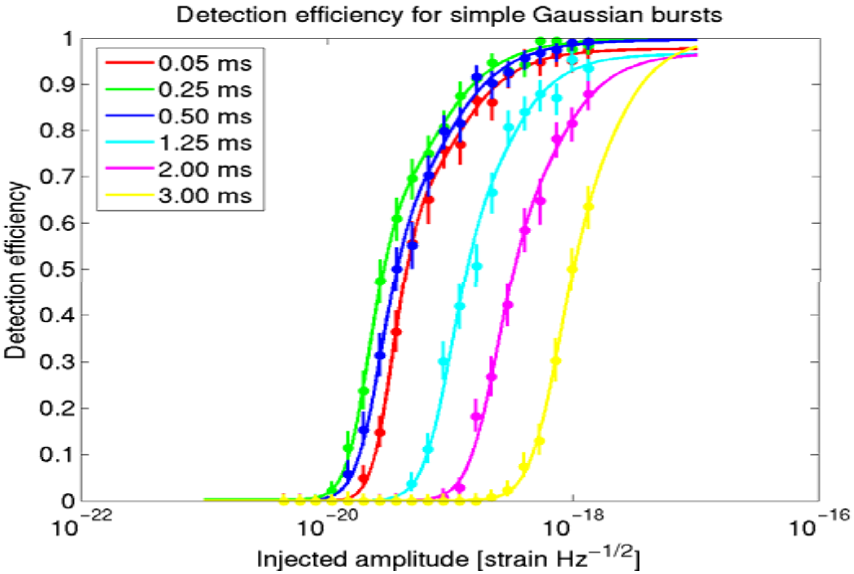


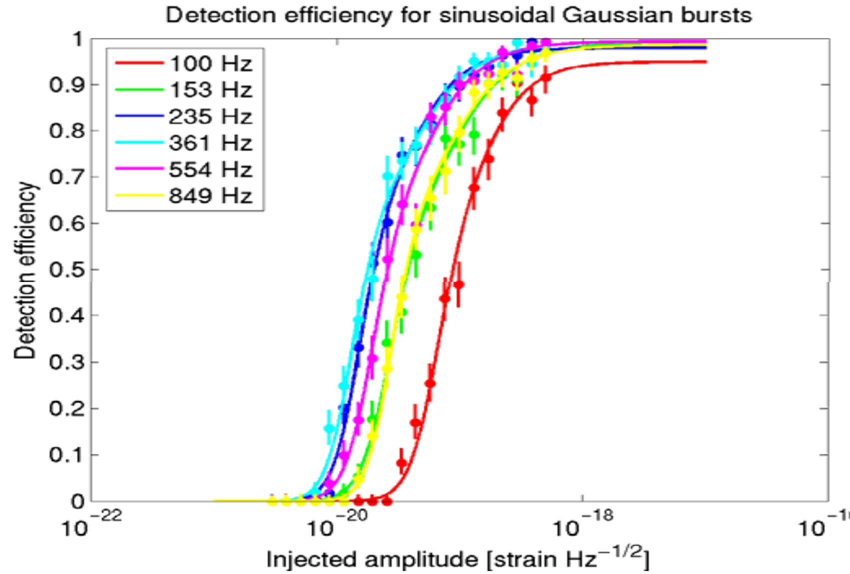


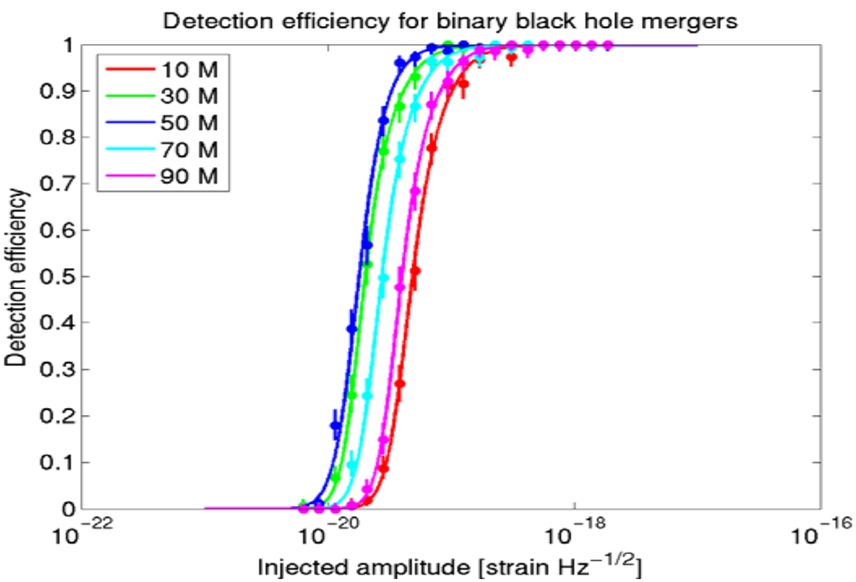


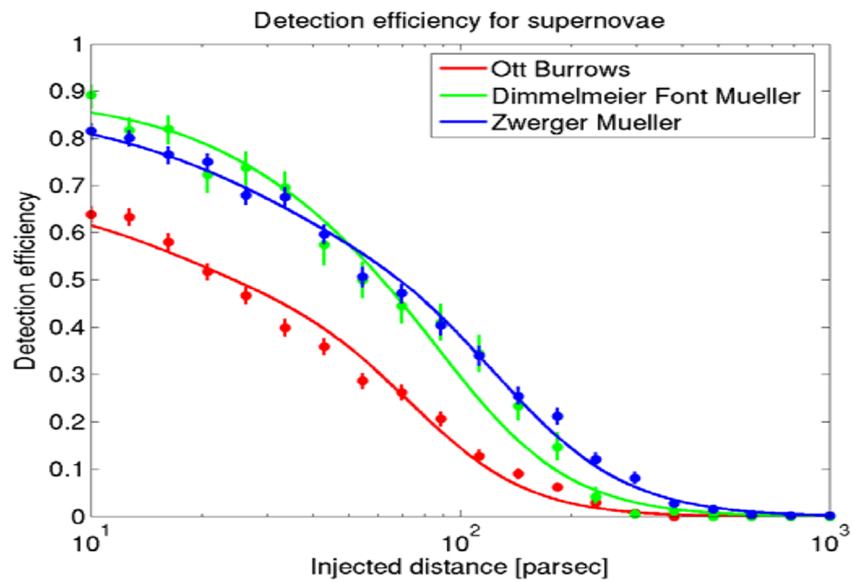


Detection efficiencies

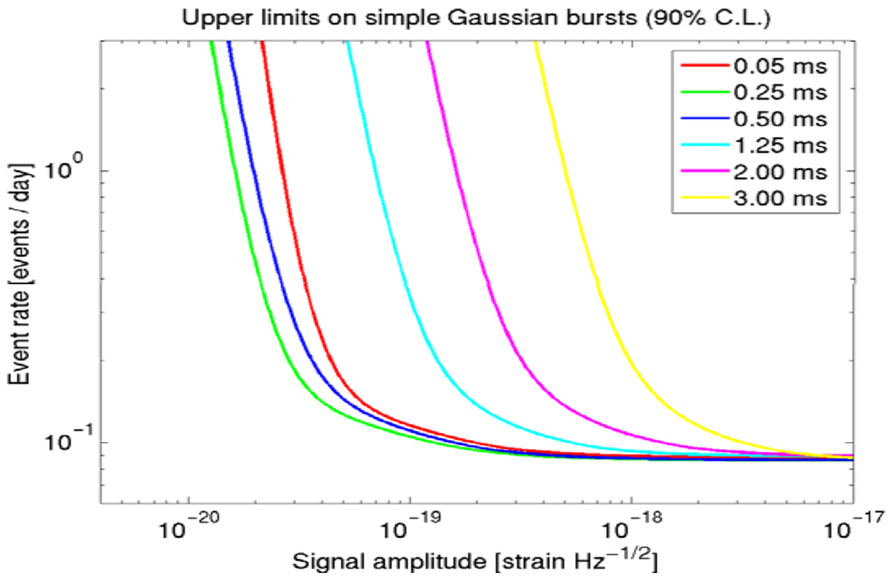


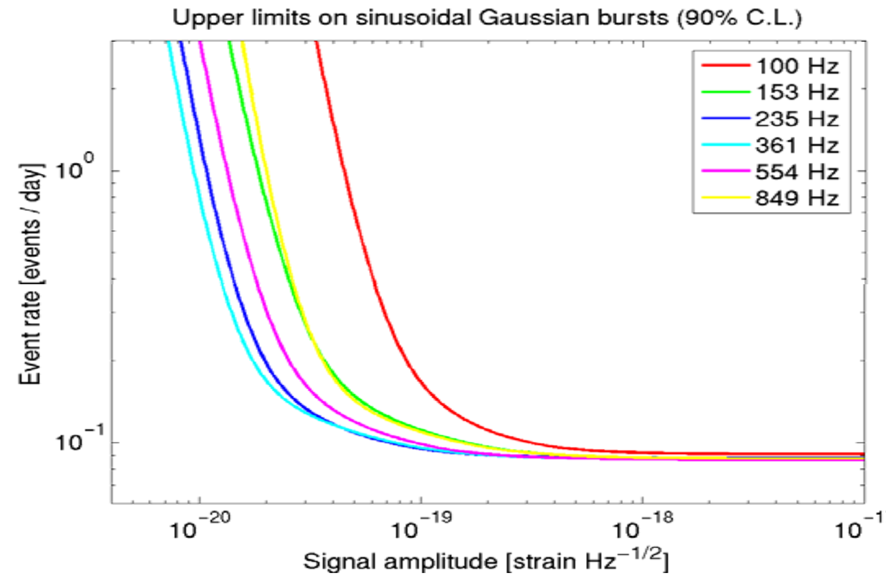


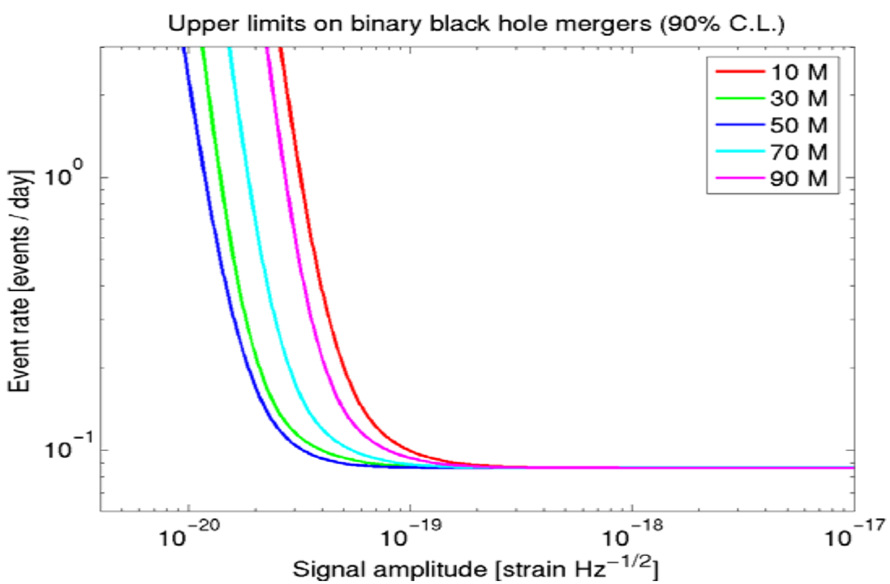


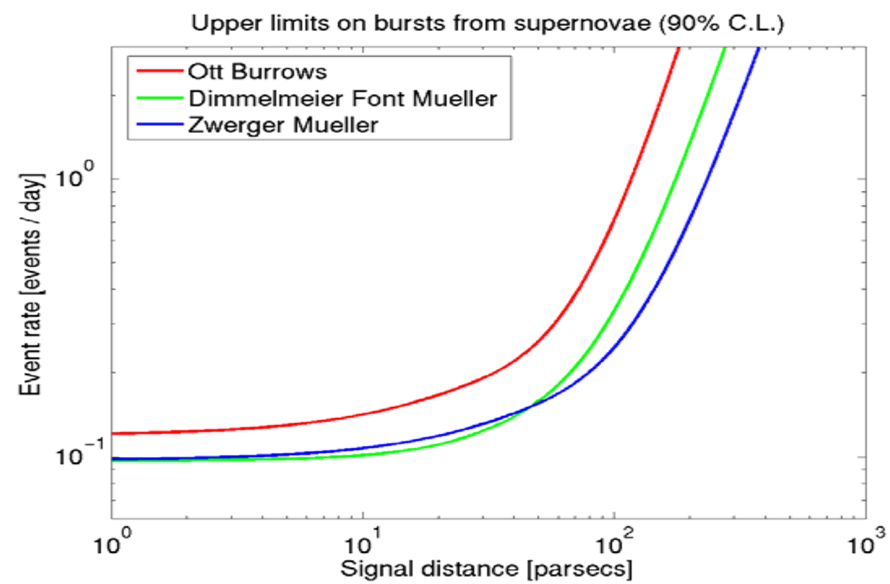


Upper limits

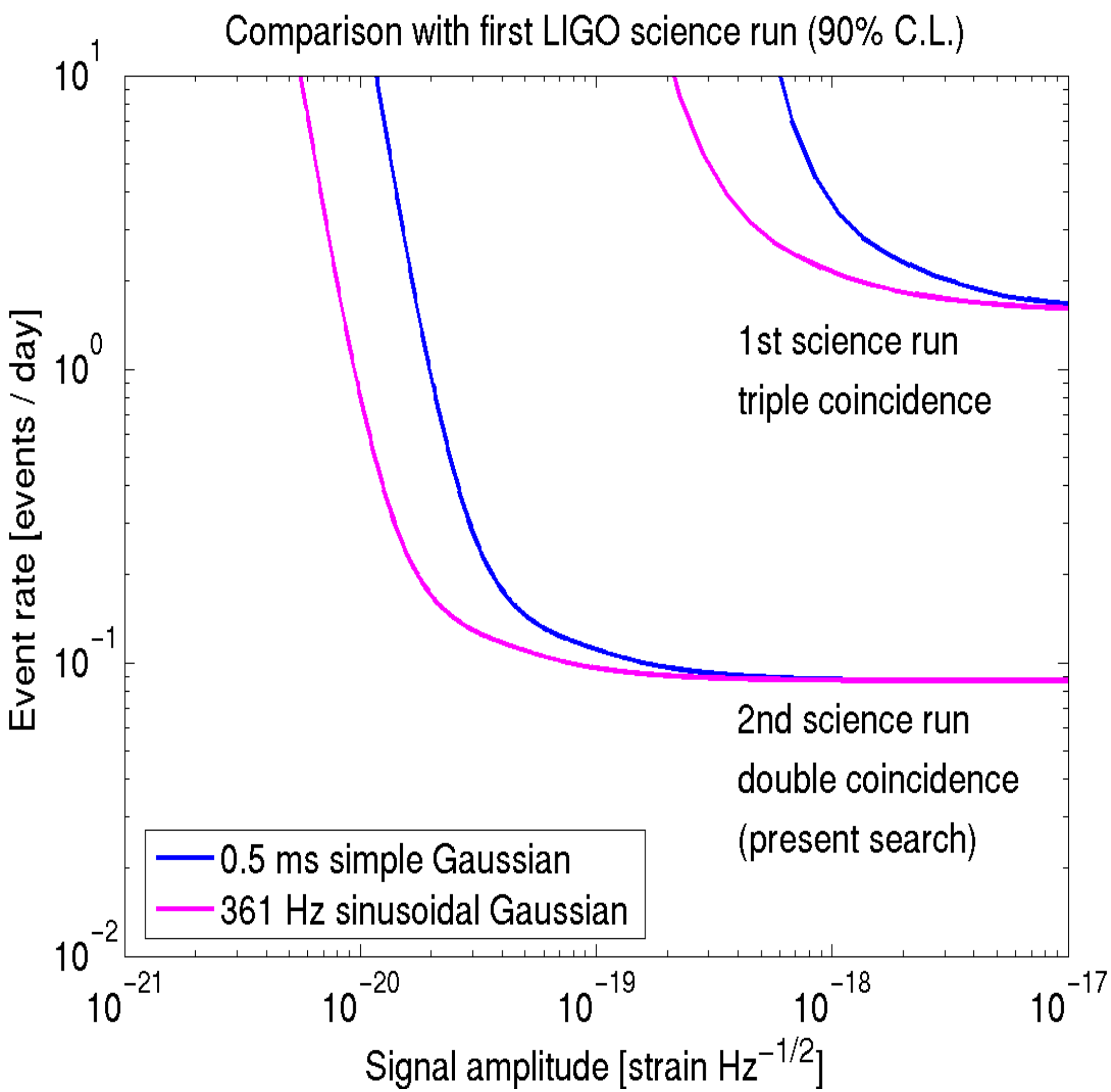






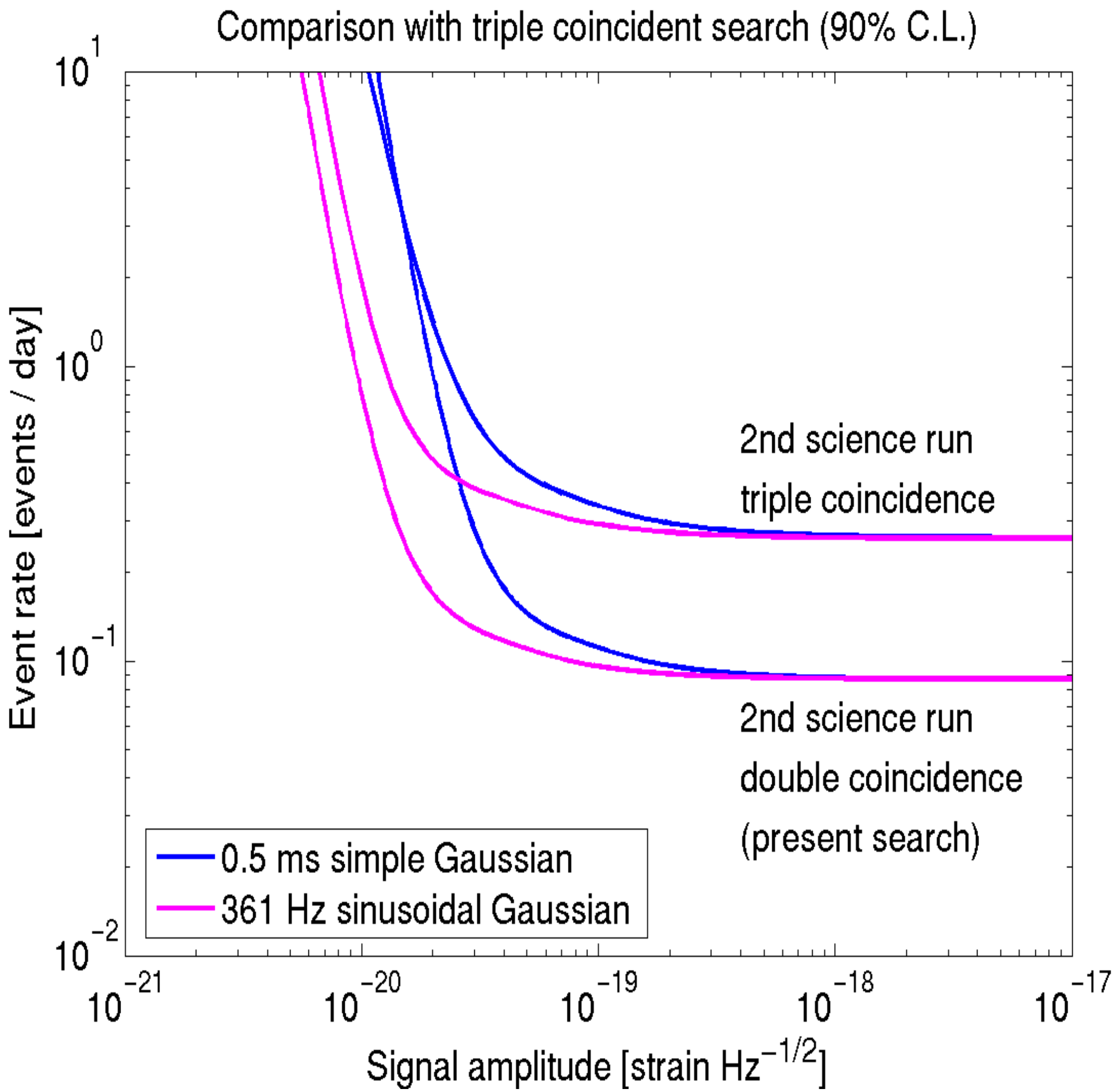


Comparison with first LIGO science run



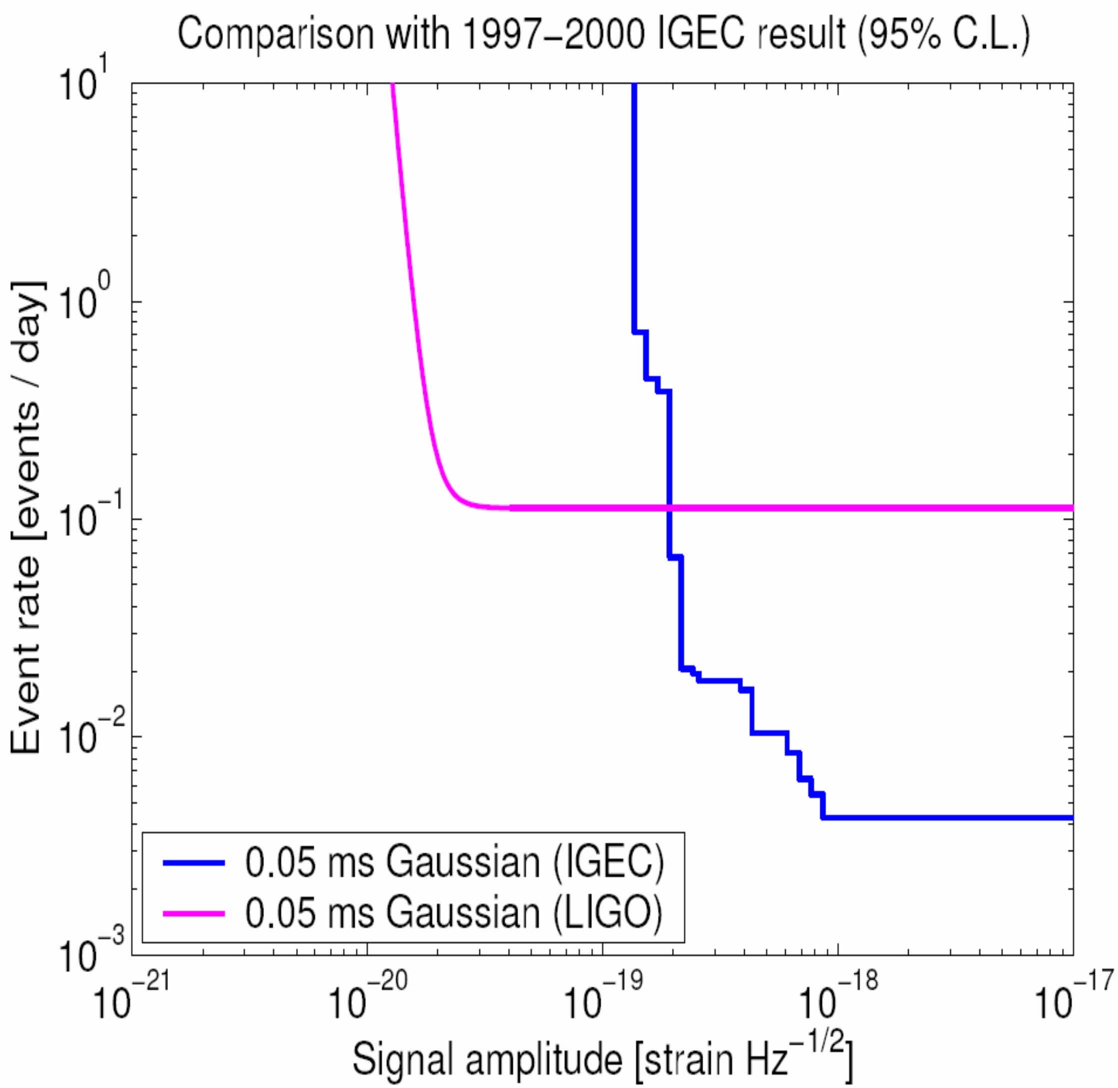
- Common set of simulated waveforms
- ~10 x improvement in detector sensitivity
- ~2 x improvement in search algorithm
- ~20 x improvement in observation time

Comparison with triple coincident search



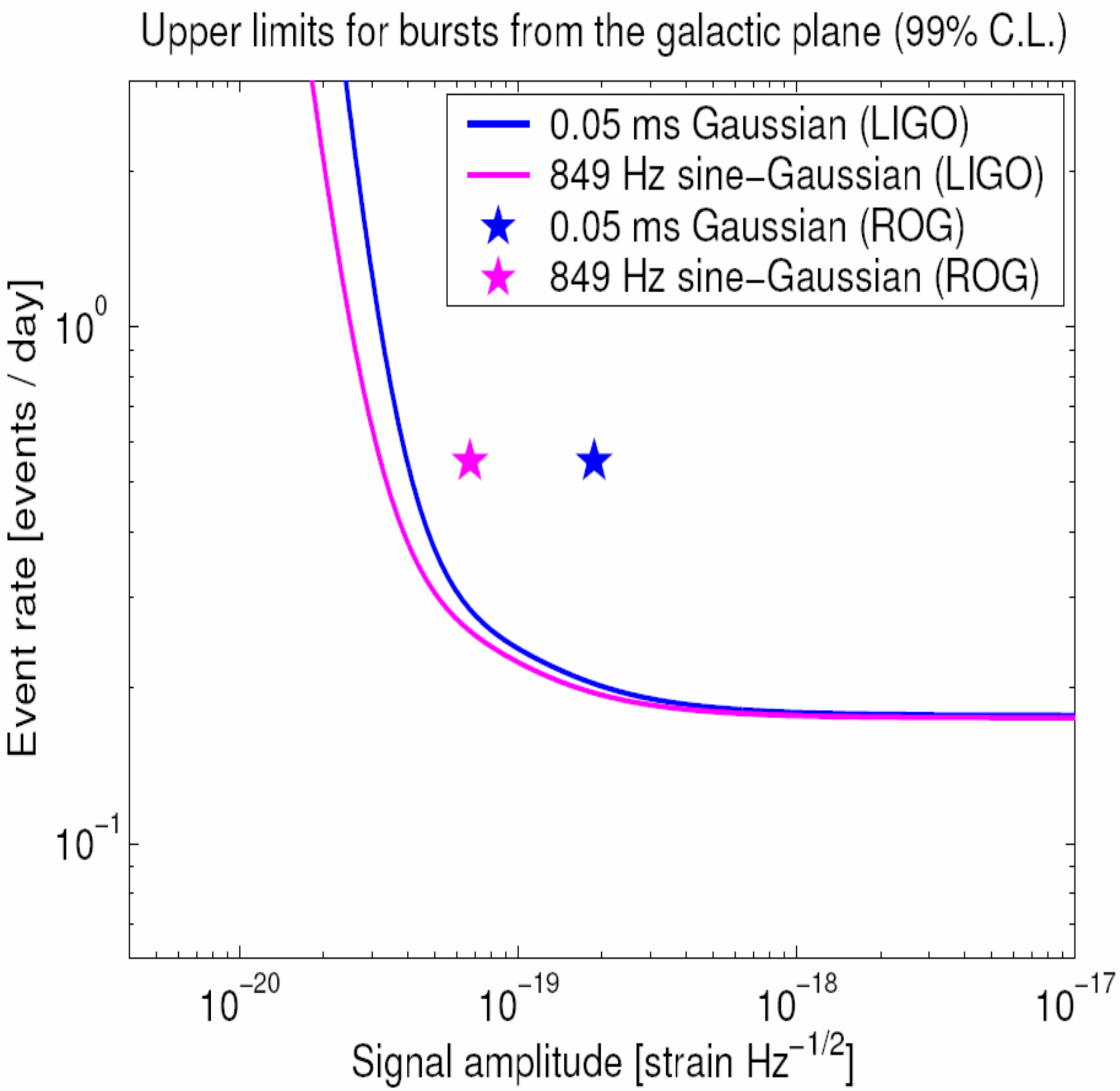
- Common set of simulated waveforms
- Similar sensitivity despite use of only two detectors
- Due to advantages of coherent search.
- ~3 x improvement in observation time

Comparison with IGEC collaboration



- Comparison is waveform specific
- Best case waveform consistent with IGEC assumptions
- IGEC search has ~33 x greater observation time
- LIGO search has ~10 x greater sensitivity for this waveform

Comparison with ROG collaboration



- Comparison is waveform specific
- Conservative choice of waveform
- Sensitivity to sources in galactic plane is similar to all sky
- ROG results are excluded at the 99% confidence level

Future prospects

- Third science run
  - Decreased acoustic, seismic, and RF coupling
  - Factor of 2 to 5 improvement in sensitivity
  - ~25 percent increase in observation time
- Fourth science run
  - Order of magnitude improvement in sensitivity
  - Within a factor of a few of design sensitivity
  - ~25 percent less observation time
- Fifth science run
  - One year of observation at design sensitivity!
  - Commencing Fall 2005!

Room for improvement

- More extensive consistency testing
- Detector specific search parameters
- Evaluate performance for non-localized bursts
- Clustering of events in the time-frequency plane
- Hierarchical search
- Waveform reconstruction and parameter estimation
- Directed search for bursts
- Detector characterization and veto studies

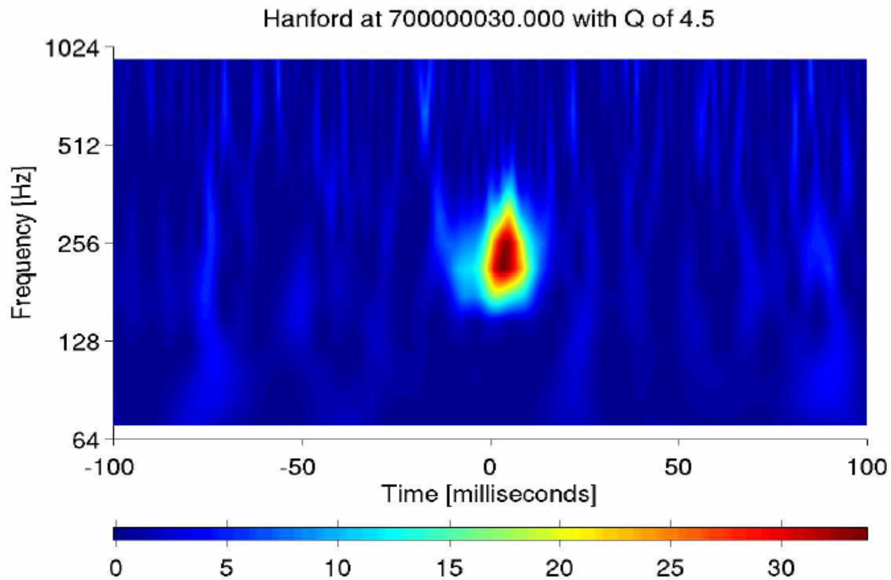
Detector characterization and vetoes

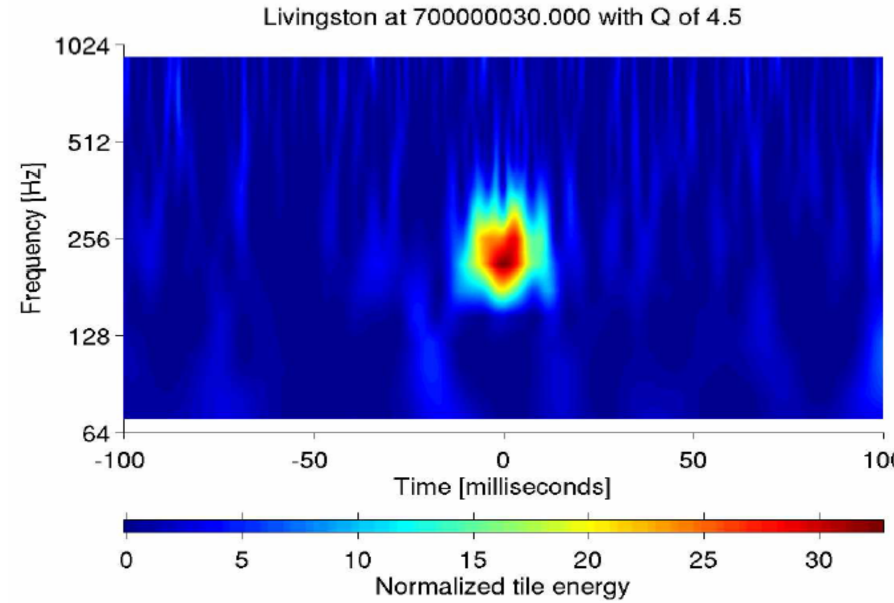
- Q Pipeline shows good prospects for detector characterization and veto investigations.
- Currently performing a full search of the Hanford level 1 reduced data set from S3 and S4 to identify potential veto channels.
- Developing a tool to post-process environmental and auxiliary interferometer channels around the time of interesting events.
- Hope to provide a set of tools for control room use during S5.

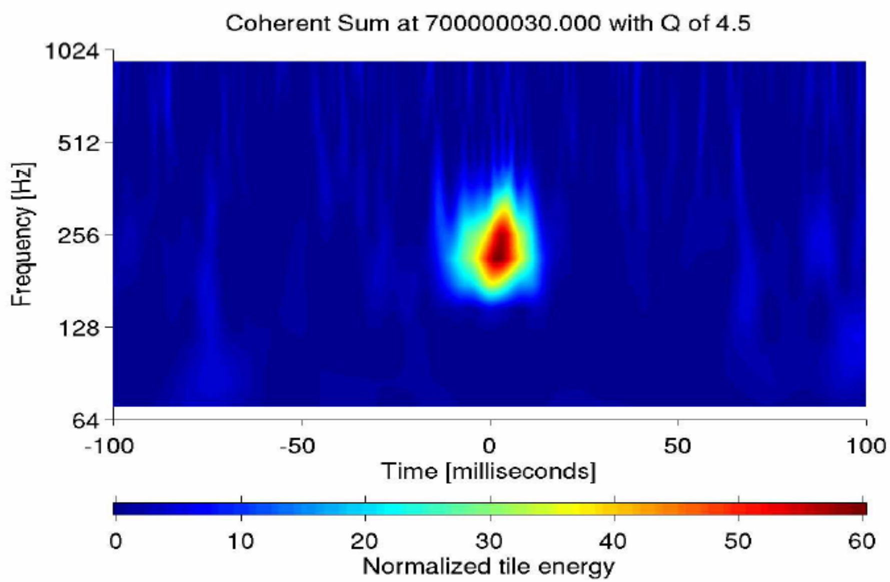
Directed search

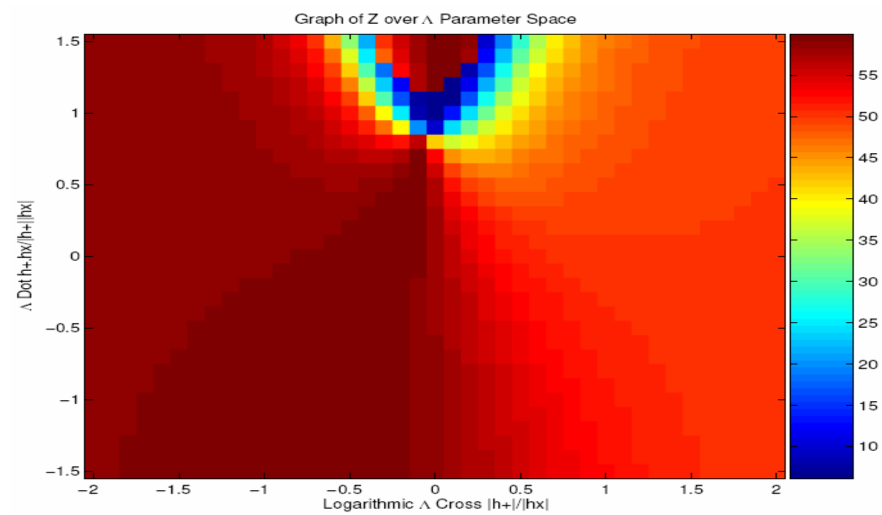
- Quasi-coherent search to target a position of interest on the sky using two or more detectors
- Design and implementation by Sahand Hormoz, University of Toronto, Caltech SURF student
- Based on method proposed by Julien Sylvestre
- Two detectors do not provide complete information
- Search over a two-dimensional parameterization of waveform space
- Maximize signal to noise ratio

Directed search example





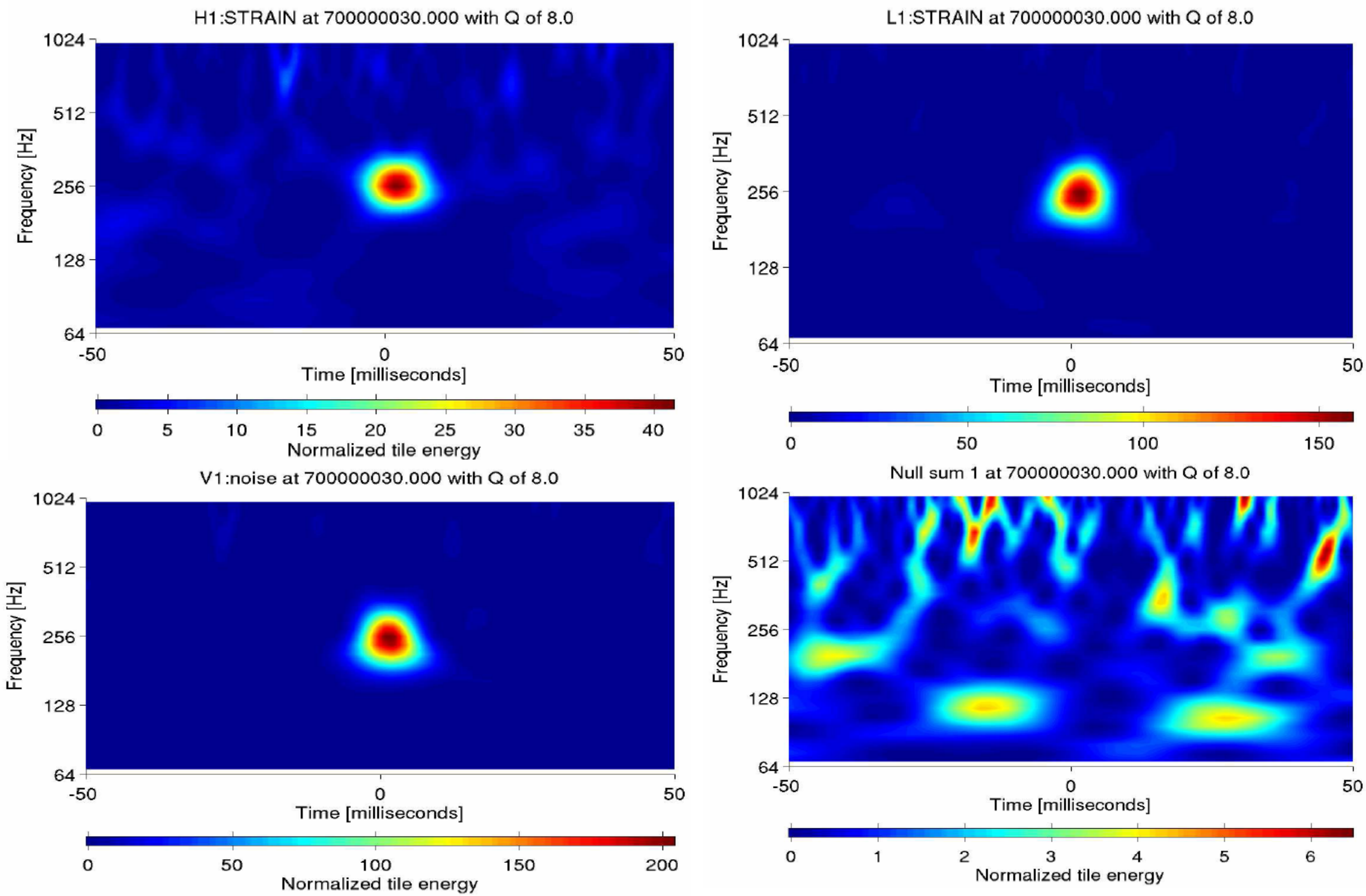




Coherent search

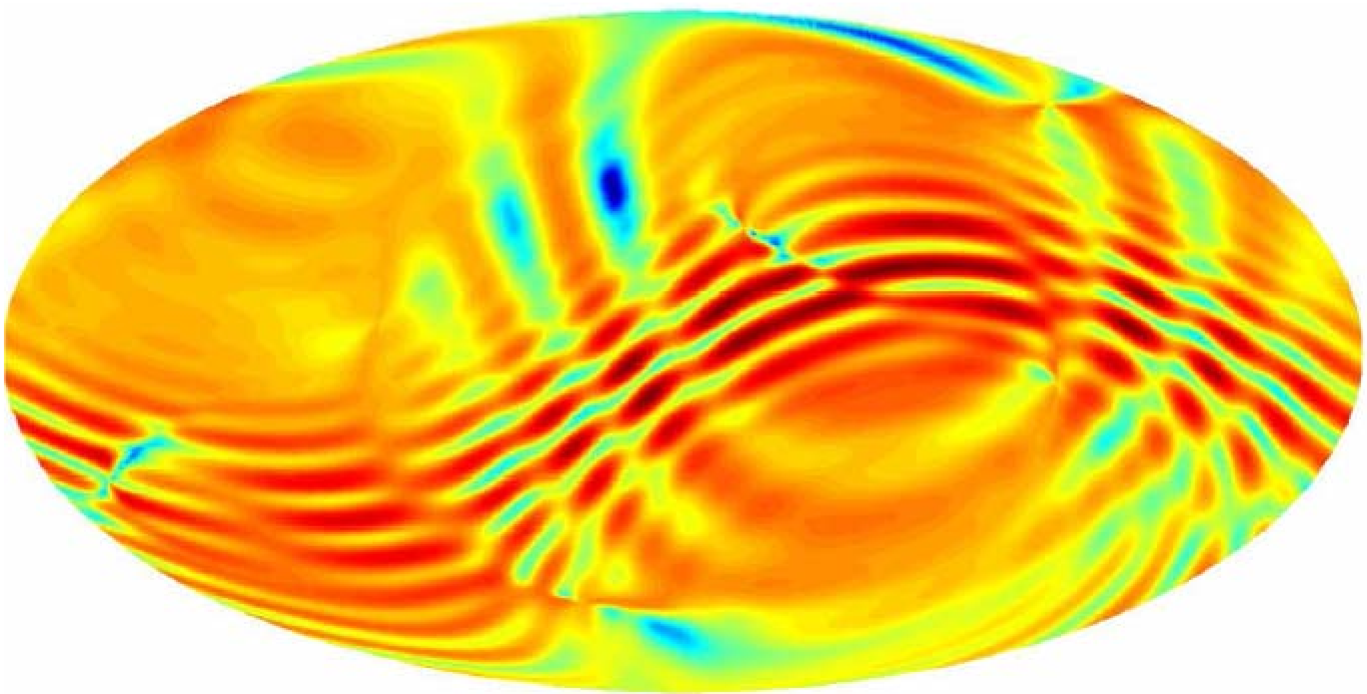
- Three or more non-aligned detectors provide sufficient information to reconstruct signal.
- Work with Albert Lazzarini, Patrick Sutton, Massimo Tinto, Antony Searle, and Leo Stein.
- Generalization of the approach of Gursel-Tinto to more than three detectors
- Construct linear combinations which cancel the signal
- Test for consistency with noise
- Search over the sky

Coherent search example



Coherent search example sky map

- Sine-Gaussian burst in simulated detector noise
- From the galactic center



Consistency testing

- Test difference between H1 and H2 transforms for consistency with detector noise
- Test difference between arbitrary detectors after accounting for detector response and a assumed sky position
- How can calibration error and/or incorrect sky position be taken into account?
- Issues being studied by Sahand Hormoz
- Intend to apply to S3 and S4 double coincident search of Hanford data