Bayesian approaches to the detection and analysis of unmodeled gravitational wave signals

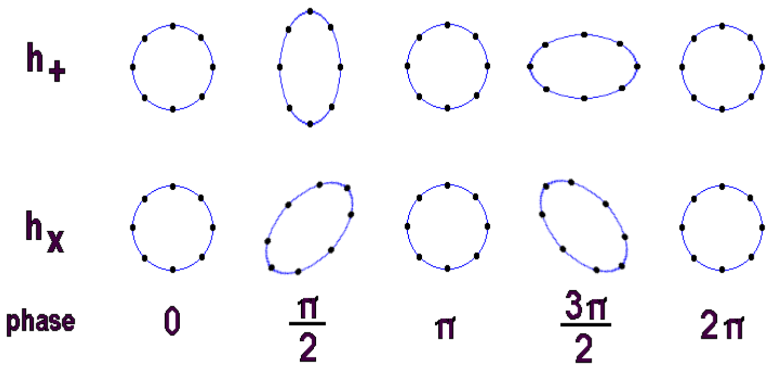
Talk by Ryan Lynch
MIT LIGO Lab
October 21 2014

Collaborators: Salvatore Vitale, Reed Essick

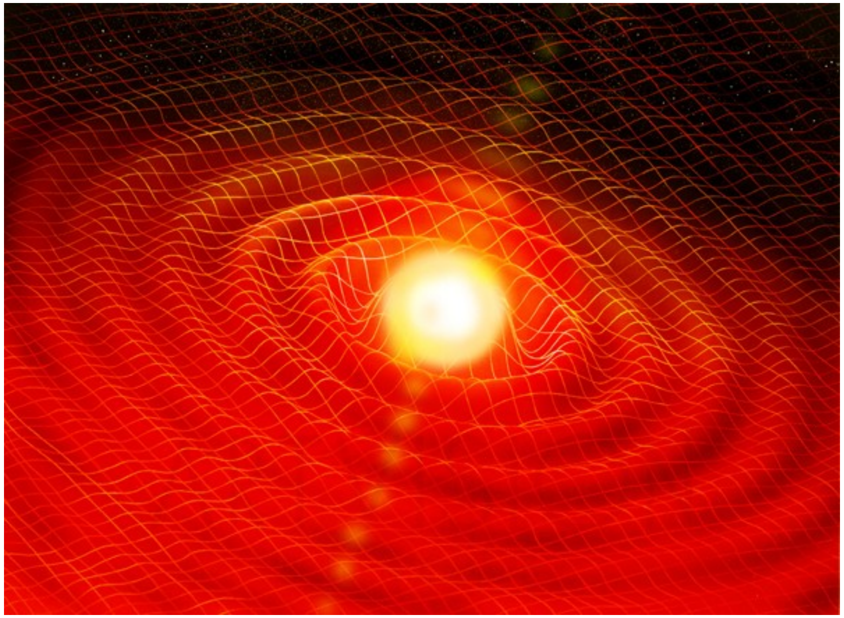
Background on LIGO

Gravitational Waves

- Caused by disturbances to a stable spacetime manifold
- Expected to propagate at speed of light
- Expected to have 2 independent polarizations
- Sources
  - **<u>Unmodeled bursts</u>**
  - Binary Coalescence
  - Periodic Sources
  - Stochastic Background



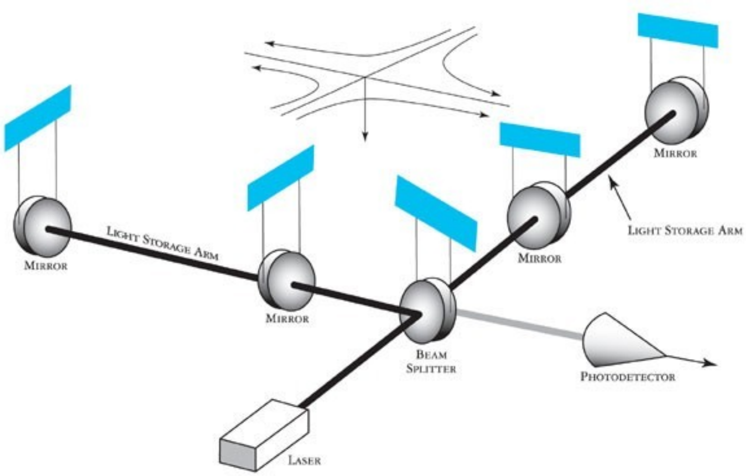
http://www.johnstonsarchive.net/relativity/pictures.html



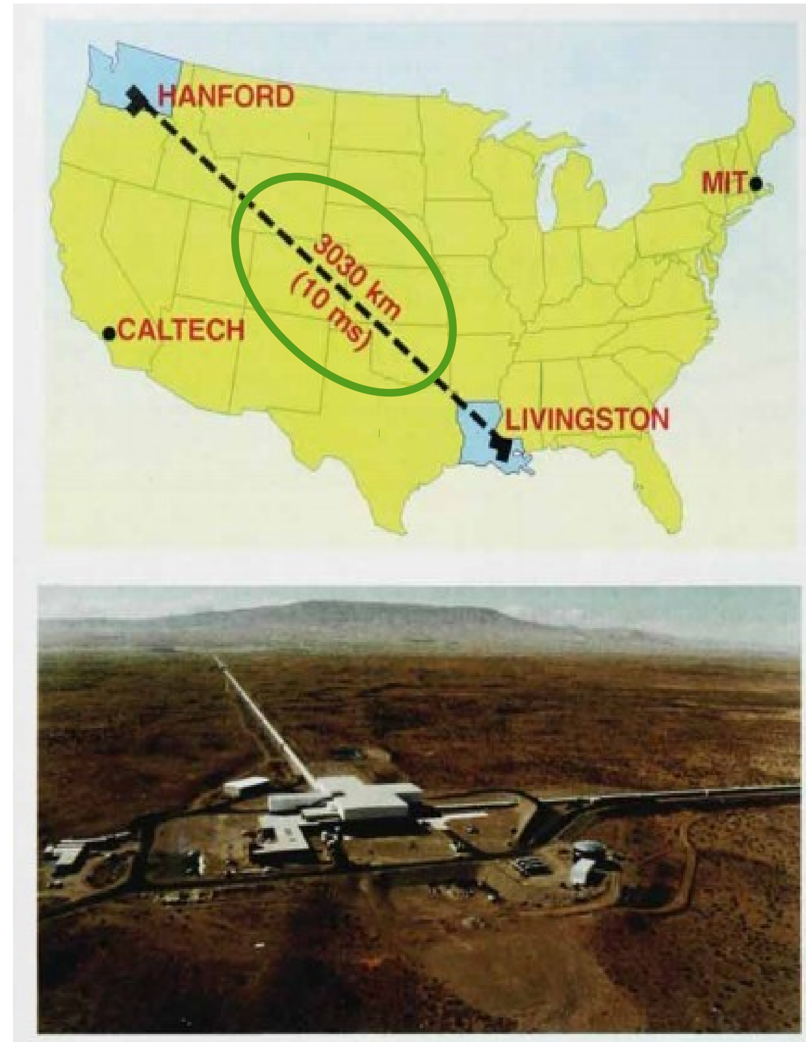
http://mucholderthen.tumblr.com/post/55970198303/gravitational-waves-patterns-in-space-time-a

LIGO-VIRGO Interferometer Network

- Observatories are large Michelson interferometers (to 0^{th} order)
- LIGO: Two observatories with 4 km arms in US (Hanford, Washington & Livingston, Louisiana)
- Virgo: One observatory with 3km arms (Cascina, Italy)



http://www.ligo.caltech.edu/LIGO_web/PR/scripts/draw_lg.html



http://phys.columbia.edu/~millis/1900/readings/LIGO.pdf

Signal Detection and Analysis

- Matched filtering
  - Assume data is of form: $d(t) = h(t) + n(t)$
    - h(t) is gravitational wave signal
    - n(t) is detector noise
  - Define noise-weighted inner product

$$(A|B) = 2\int_0^\infty df\,\frac{\tilde{A}^*(f)\tilde{B}(f)+\tilde{A}(f)\tilde{B}^*(f)}{S(f)}$$ where $$S(f) = 2 < \tilde{n}(f)\tilde{n}^*(f) >$$

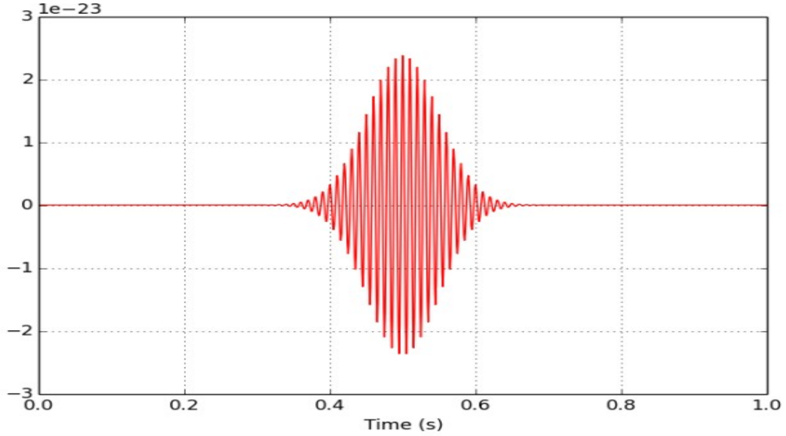
- Assuming stationary Gaussian noise, can write likelihood (under hypothesis H) as

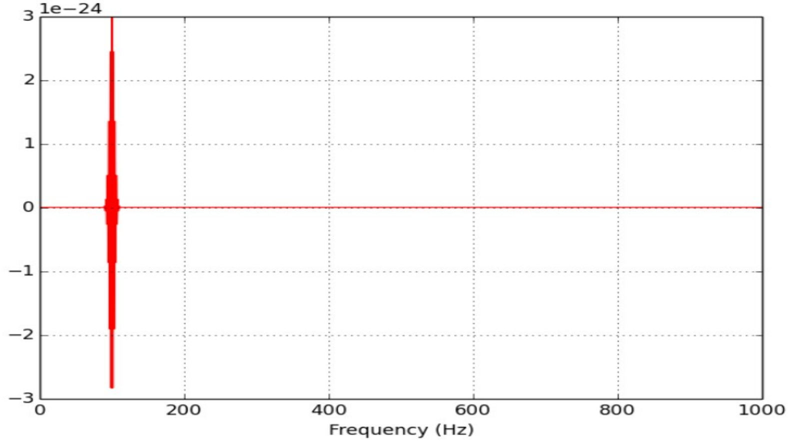
$$p(d|h,H) \propto \exp[-\tfrac{1}{2}(d-h|d-h)]$$

MCMC Approach

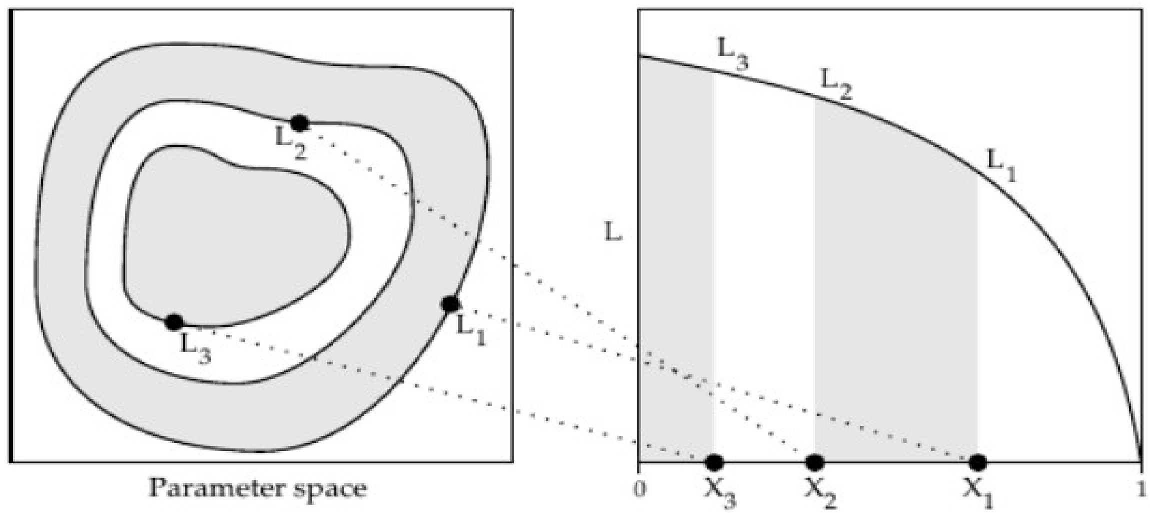
LALInference Burst

- Used MCMC-based nested sampling to explore parameter space
  - Uses a Sine-Gaussian template to model waveform
  - Calculates evidence Z directly
  - Parameter posteriors can be obtained by resampling MCMC points





Nested Sampling (Skilling 2004)



Skilling 2004

- Reparametrize evidence $Z = \int L(\vec{\theta})\, p(\vec{\theta})\, d\vec{\theta}$ with $X(\lambda) = \int_{L(\vec{\theta})>\lambda} p(\vec{\theta})\, d\vec{\theta}$

$$Z = \int L(X)\, dX \approx \sum_i L(X_i) \Delta X_i$$

- Scatter "live points" through initial Monte Carlo (i.e., draw $\vec{\theta}$ from prior)
- Calculate likelihood at each live point
- By drawing from prior, replace lowest likelihood live point with new live point
  - New point of higher likelihood: forms nested likelihood contours
- Algorithm calculates Z while converging to max likelihood

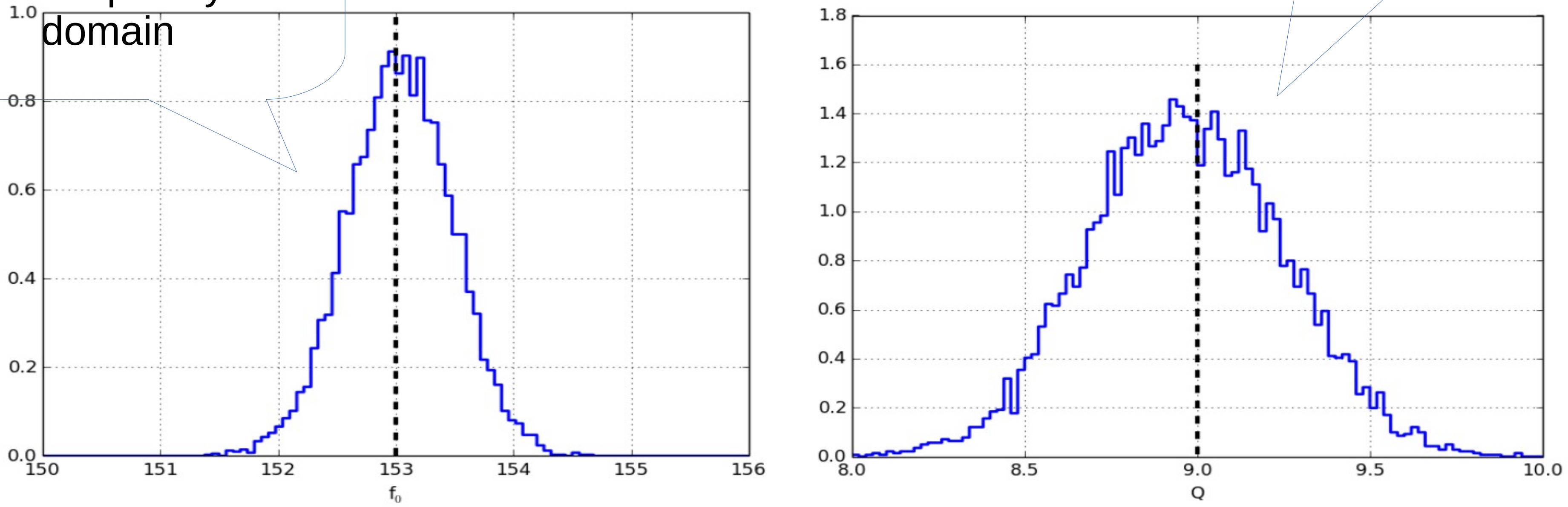
Parameter estimation and posteriors

- If live points are recorded, have access to the posterior through

f_0 is oscillation frequency in t-domin, central frequency in f-domain

$$p(\vec{\theta}_i|d, H) \propto p(\vec{\theta}_i|\mathrm{NS})\, p(d|\vec{\theta}_i, H)\, X_i$$

Q parametrizes number of waveform cycles



Courtesy of R. Essick

Bayes Factors and Signal Detection

- Write Bayes' theorem as: $p(H_i|\{d\}) = \frac{p(\{d\}|H_i)\,p(H_i)}{p(\{d\})}$
- Taking the "odds ratio" of two hypotheses we find the important quantity to be the Bayes factor:

$$B_{i,j} \equiv \frac{p(\{d\}|H_i)}{p(\{d\}|H_j)}$$

where

$$Z_i \equiv p(\{d\}|H_i) = \int_{\vec{\theta}} p(\{d\}|H_i, \vec{\theta})\, p(\vec{\theta}|H_i)\, d\vec{\theta}$$

- Signal vs. Gaussian Noise:

$$\mathcal{L}(h) \propto \exp[-\frac{1}{2}(d-h|d-h)]$$

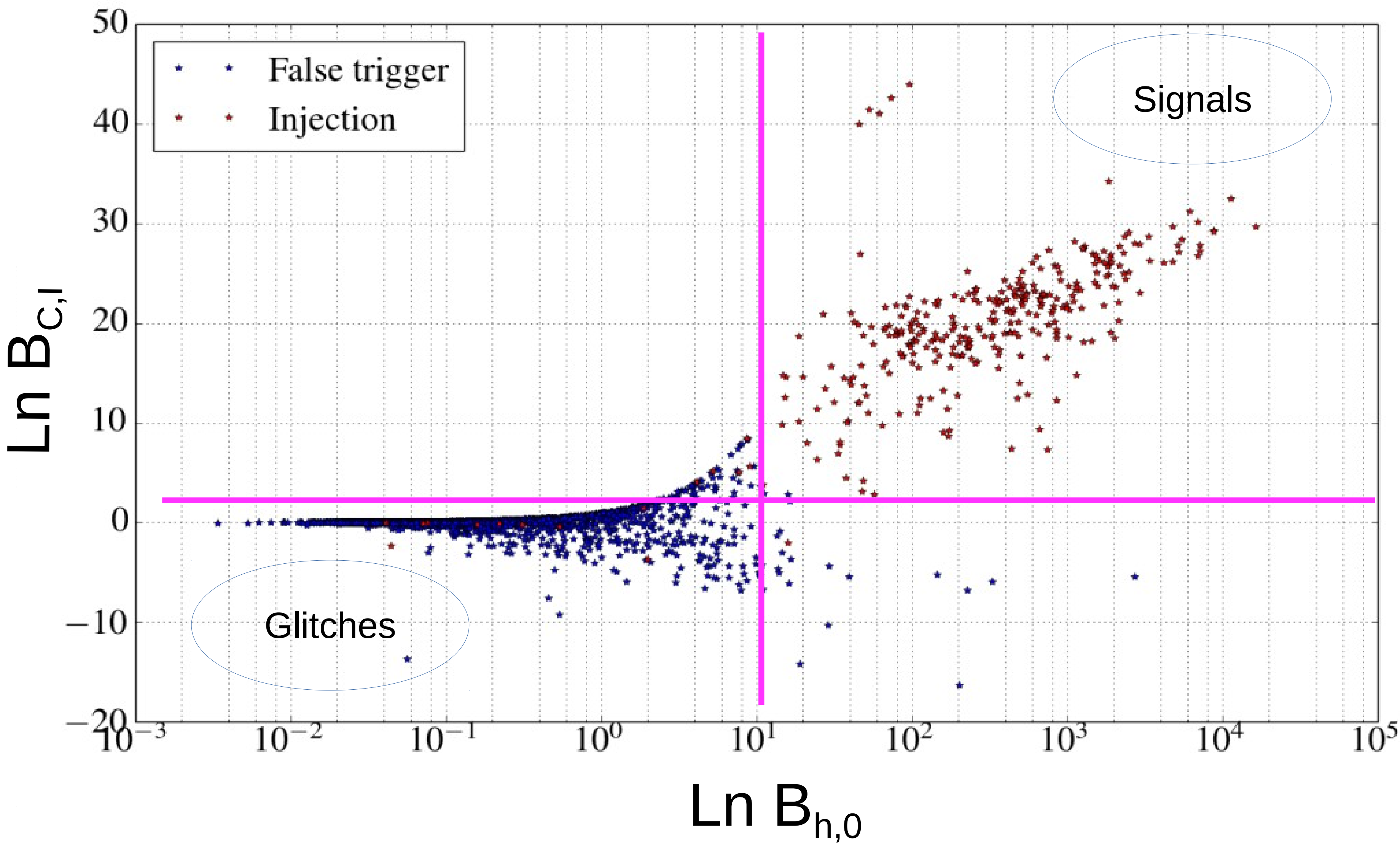
$$B_{h,0} = \frac{Z_h}{Z_0} = \frac{\int \mathcal{L}(h(\vec{\theta}))\, p(\vec{\theta})\, \mathrm{d}(\vec{\theta})}{\mathcal{L}(h=0)}$$

- Coherent Signal vs. Incoherent Glitches:

Useful because coherent prior is more sharply peaked in parameter space

$$B_{C,I} = \frac{Z_{\{\beta\}}}{\prod_\beta Z_\beta} = \frac{\int \mathcal{L}(\{h_\beta(\vec{\theta})\})\, p(\vec{\theta})\, \mathrm{d}(\vec{\theta})}{\prod_\beta \int \mathcal{L}(h_\beta(\vec{\theta}_\beta))\, p(\vec{\theta}_\beta)\, \mathrm{d}(\vec{\theta}_\beta)}$$

Signals vs. False Alarms

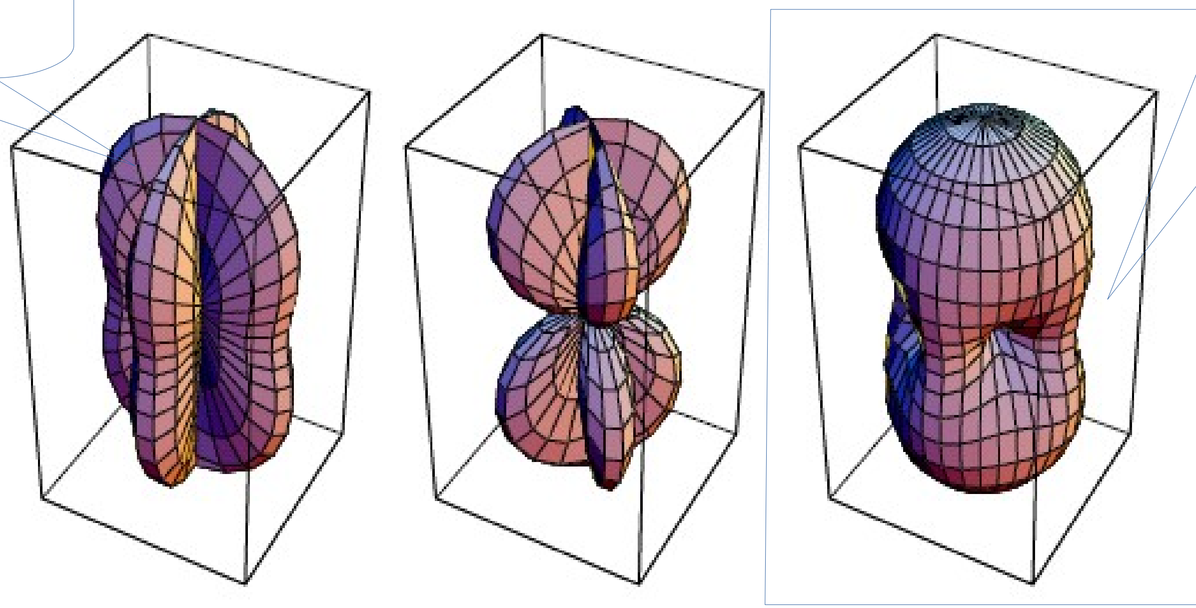


Low-latency sky localization

Single Detector Sensitivity

Most sensitive when source is overhead

“Null” spots when source is in plane of detector



http://docuserv.ligo.caltech.edu/docs/public/T/T970101-B.pdf

$F_\times$ F_+ $\sqrt{F_+^2 + F_\times^2}$

Effective polarization to which single detector is sensitive

Single detectors have very poor sky localization capabilities

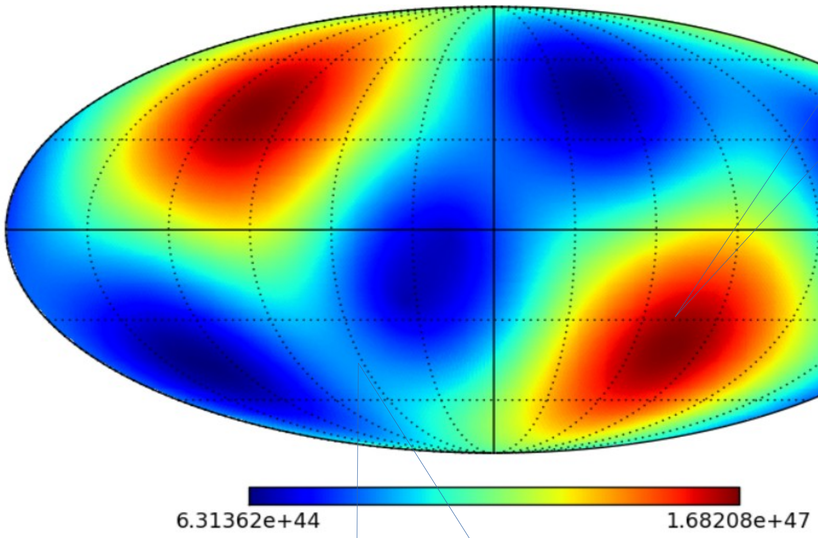
Network Sensitivity

- For 2-detector (HL) case:

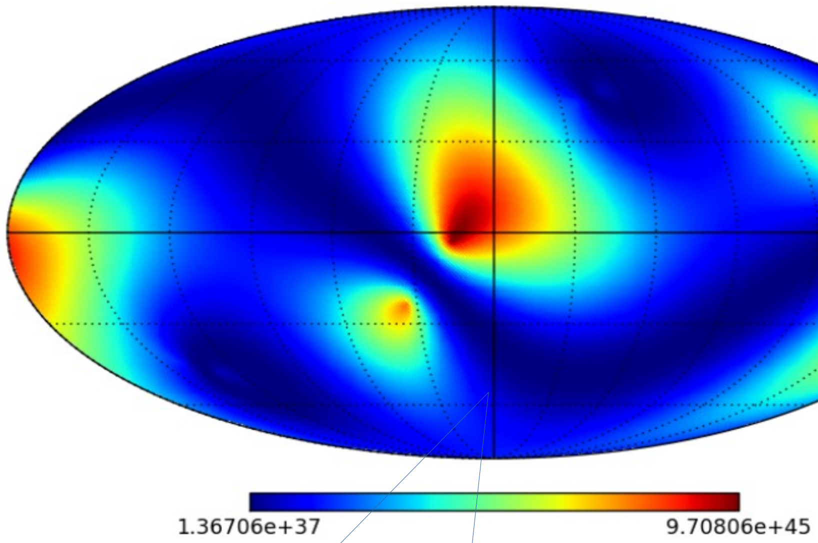
$$A_{kj} \equiv \sum_{\beta} \frac{F^*_{\beta k} F_{\beta j}}{S_\beta}$$

Motivation for priors!

Max Sensitivity Eigenvalue



Min Sensitivity Eigenvalue

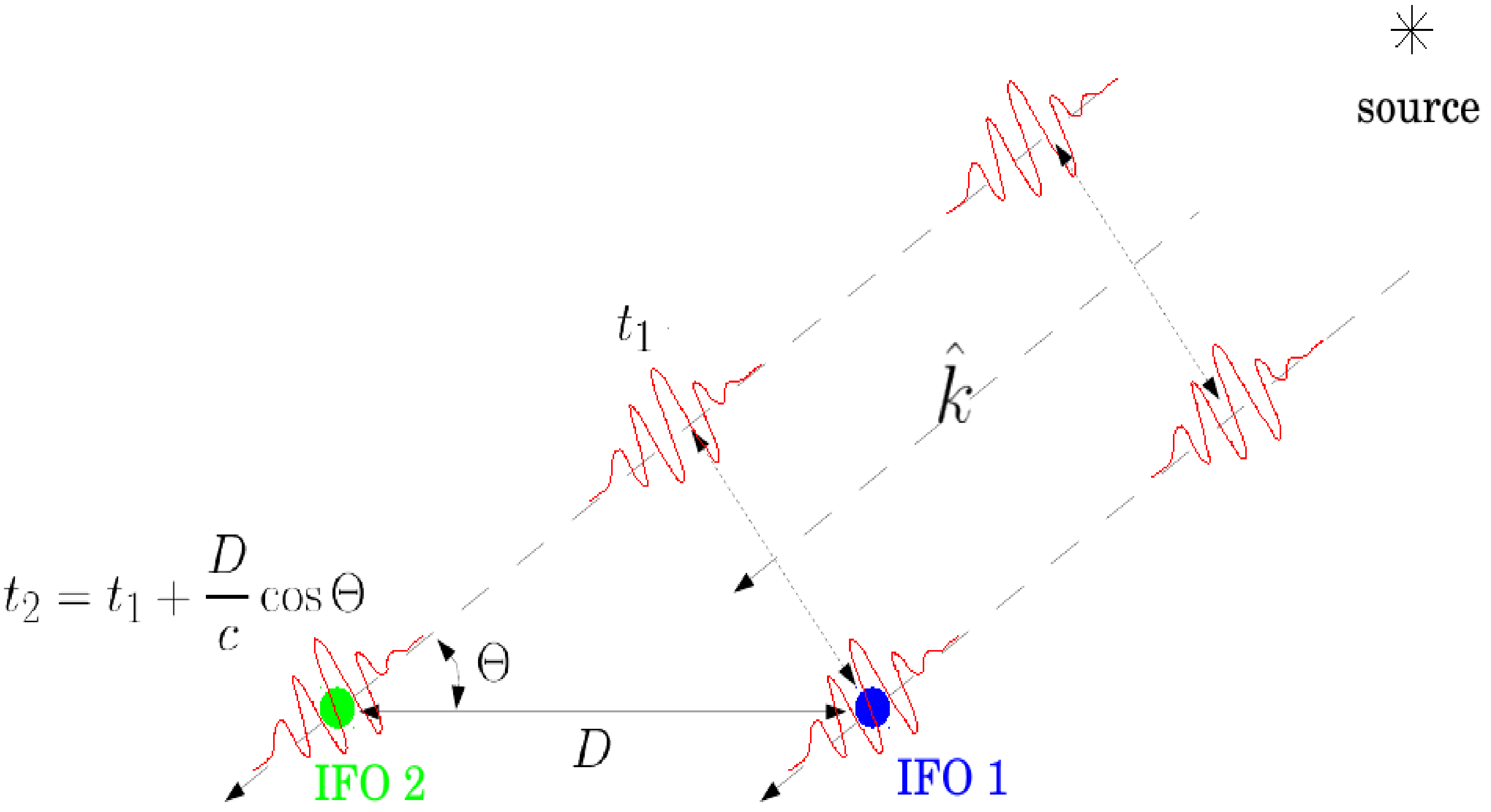


Max eigenvalue similar to sensitivity of single detector

Min eigenvalue much smaller in magnitude across most of sky

Performing Sky Localization

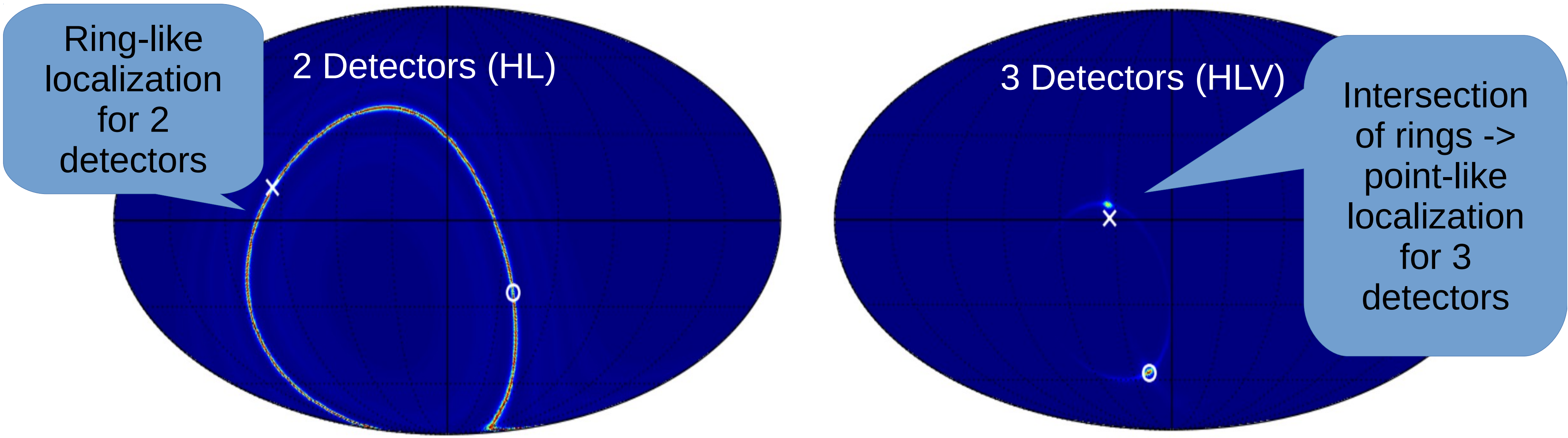
– Time-of-arrival measurements give rings on sky for each pair of detectors



Courtesy of R. Essick

Performing Sky Localization

- Triangulation:

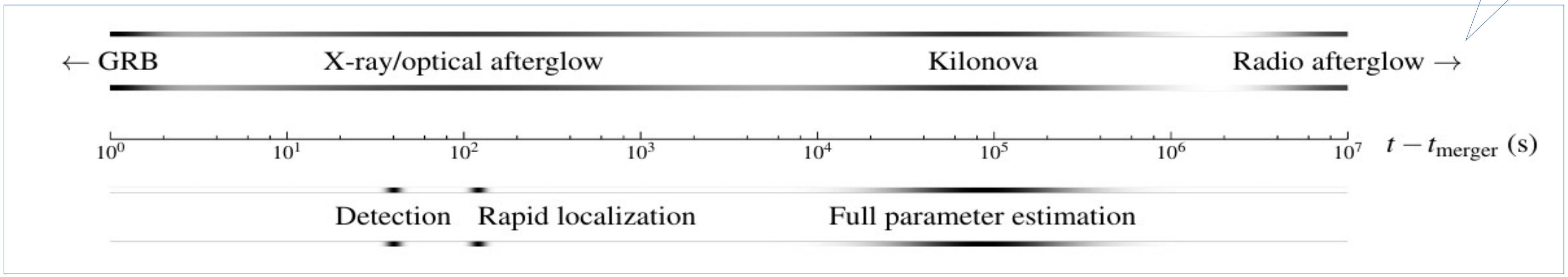


Courtesy of R. Essick

- Priors and amplitude consistency can provide modulation along timing rings

Burst Sky Localization

For BNS CBC events



http://arxiv.org/abs/1404.5623

- Some gravitational wave signals expected to have EM counterparts and afterglows
- Challenge: Need accurate localization with low latency
  - Searching over entire sky is computationally expensive
  - Sky location posteriors tend to be fragmented and non-localized
  - How to model burst signals and search over parameter space?

Low-latency Bayesian Approach

- Can do full parameter estimation and sky localization follow-up with LIB on timescales of hours
- Our goal: design a low-latency, all-sky sky localization pipeline
  - Allow for varying degree of signal strain (h(f)) modeling
    - Marginalize over all strain amplitudes through Gaussian integration
    - This requires expansion of prior in terms of Gaussians
  - Make search coherent among detectors: enables amplitude consistency

Likelihood

Ratio of Gaussian noise realizations with and without signal present

Beta signifies the detector, i,j signify the polarization

- Define Likelihood ratio as:

$$\mathcal{L} = \frac{(d_\beta - F_{\beta j} h_j e^{-2\pi i f t_0}|0)}{p(d_\beta|0)} = \exp\left(\frac{2}{T}\sum_{f,\beta}\frac{|d_\beta|^2 - \left|d_\beta - F_{\beta j} h_j e^{-2\pi i f t_0}\right|^2}{S_\beta}\right)$$

  - d(f) is the detected data
  - h_j (f) is gravitational wave strain (j^{th} polarization)
  - $F_{x,+}(\theta, \Phi)$ are antennae patterns
  - t_0 is signal's central time
  - S(f) is noise PSD

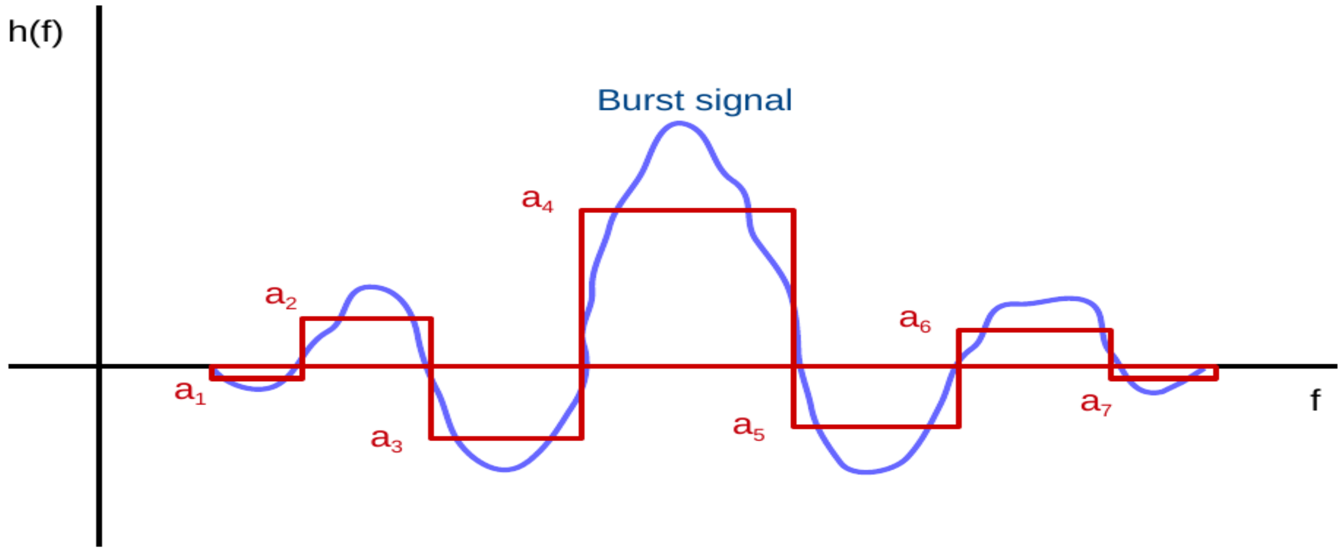
- Expansion reveals useful a quantity to be:

Defined for each sky pixel

  - Sensitivity matrix: $$A_{kj} \equiv \sum_\beta \frac{F^*_{\beta k} F_{\beta j}}{S_\beta}$$

Strain Model

- Model strain as independent “rectangular” functions over specified frequency intervals
  - $h_i(f) = a_i$ for $f_1 < f < f_2$, else 0
- In limit $N_{intervals} \rightarrow 1$, we get a “rectangular” template
- In limit $N_{intervals} \rightarrow N_{freq\ bins}$, we have a completely unmodeled signal



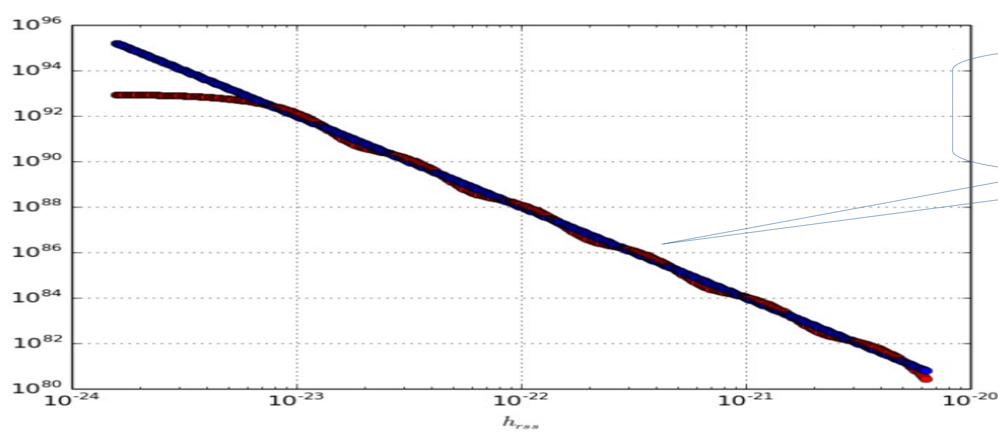
Prior on strain

- For narrow-band signals with sources uniform in volume:
  - Energy flux: $$\frac{E_{GW}}{D_L^2} \sim h_{rss}^2 \equiv \int \mathrm{d}f\, h_j^* h_j$$
  - Marginalize over energy and distance:

$$p(h_{rss}, E, D)\mathrm{d}h_{rss}\mathrm{dEdD} \propto \delta\left(h_{rss} - \sqrt{\frac{E}{D^2}}\right)\mathrm{d}h_{rss}\cdot \mathrm{d}E \cdot D^2\mathrm{d}D$$

$$\Rightarrow \quad p(h_{rss}) \propto h_{rss}^{-4} = \sum_N C_N\, e^{-h_{rss}^2 \times Z^{(N)}}$$

- Find best fits of coefficients for Gaussian expansion:



Blue is h_{rss}^{-4}
Red is Gaussian model

Final Formulation

Not necessarily true!

- Assume $p(h_{rss}) = p(\{h_i\})$
- Marginalize over each $h_j(f)$ by performing Gaussian integral:

$$p(\theta, \phi, t_0|d) \propto p(\theta, \phi)\ p(t_0) \sum_N C_N \prod_{f_{\rm ints}} \exp\left[\frac{2}{T} \times \left(\sum_f d_\beta F^*_{\beta j} e^{2\pi i f t_0}\right)^* \left(\sum_f A + Z^{(N)}\right)^{-1}_{jk} \left(\sum_f d_\beta F^*_{\beta j} e^{2\pi i f t_0}\right)\right]$$

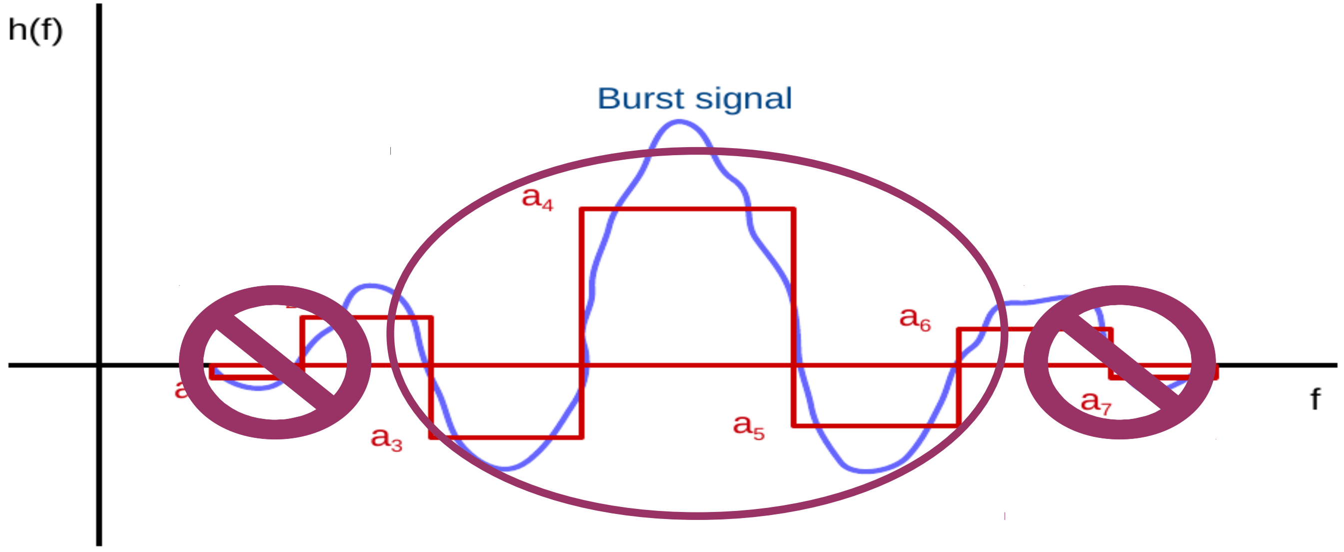
$$\times \left(\frac{\pi T}{2}\right)^{N_{\rm pol}} \left|\left(\sum_f A + Z^{(N)}\right)^{-1}\right|$$

Determinant term

- Can marginalize over t_0 using discrete fast Fourier transforms
- Dilemmas:
  - In “unmodeled” limit: determinant term acts as Occam factor that penalizes us for overfitting the data
  - In single “rectangular” limit: don't want to include frequency bins without signal

Model selection

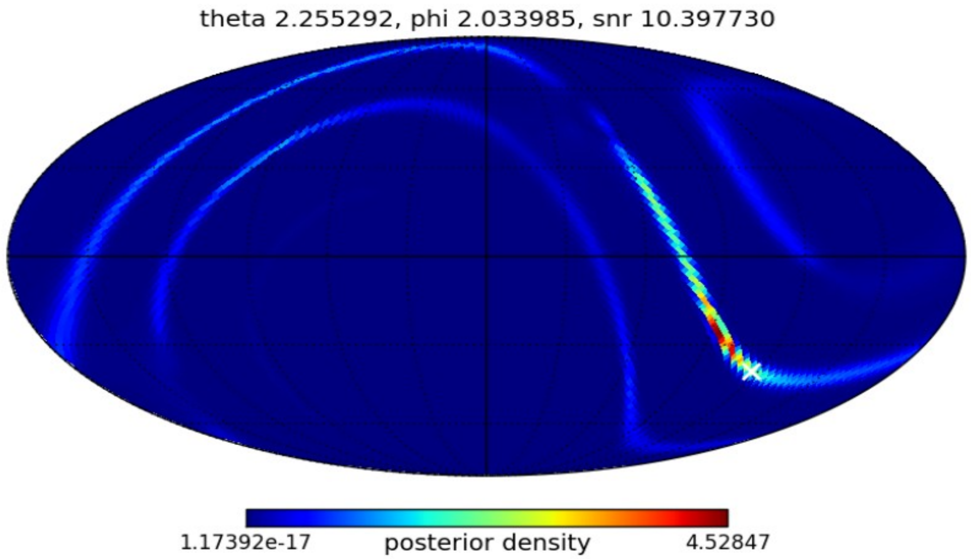
- Integrate over sky position to get a Bayes factor for signal vs. noise
- Can use prior to set h(f) to zero at any frequency bin
  - Creates “window” where signal is allowed to live
  - Reduces number of parameters
- The proper thing to do would be to marginalize over a grid of “window” models
- In favor of computational speed, currently just maximize over a set of “windows” designed to converge on true signal



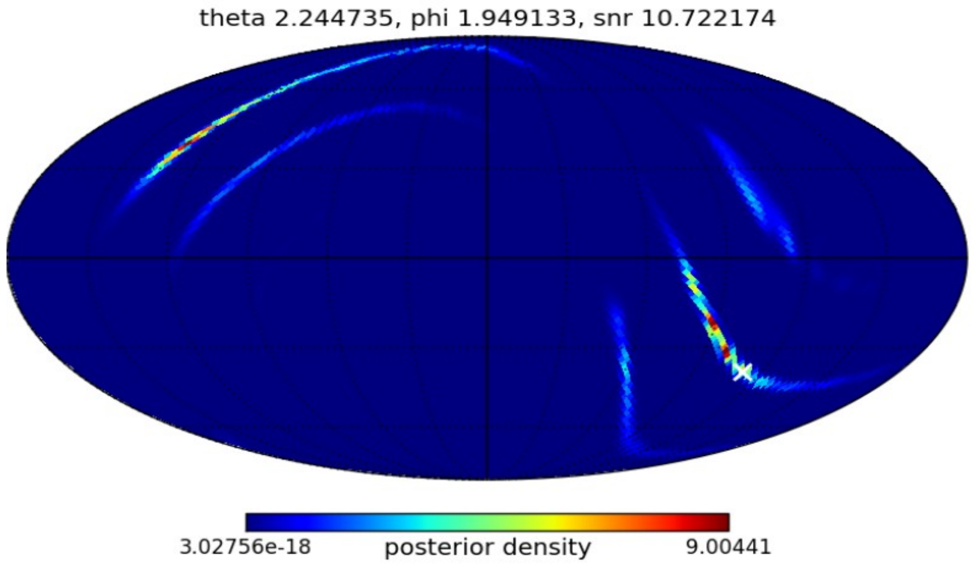
Preliminary Results ("Rectangular")

- Threshold events:

2-detector (HL) network



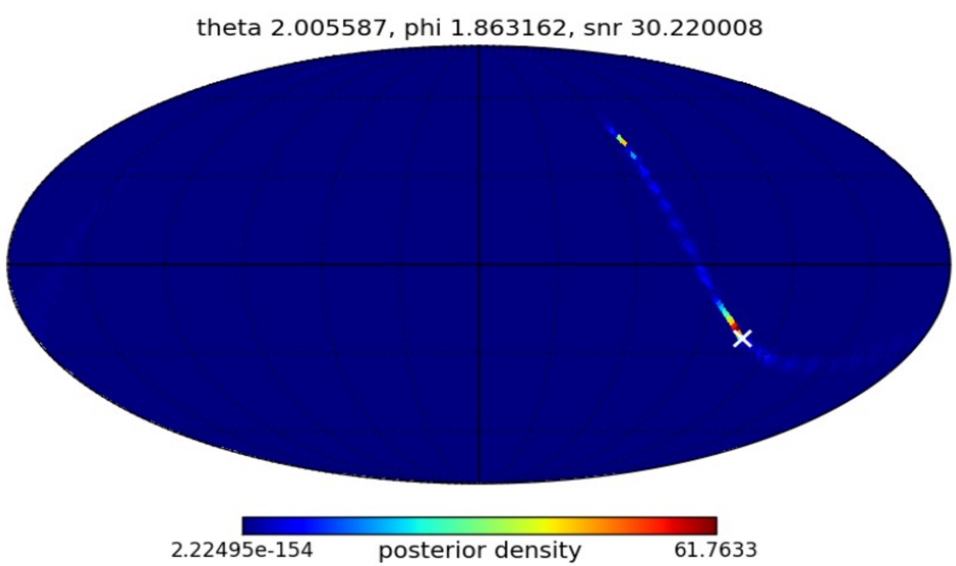
3-detector (HLV) network



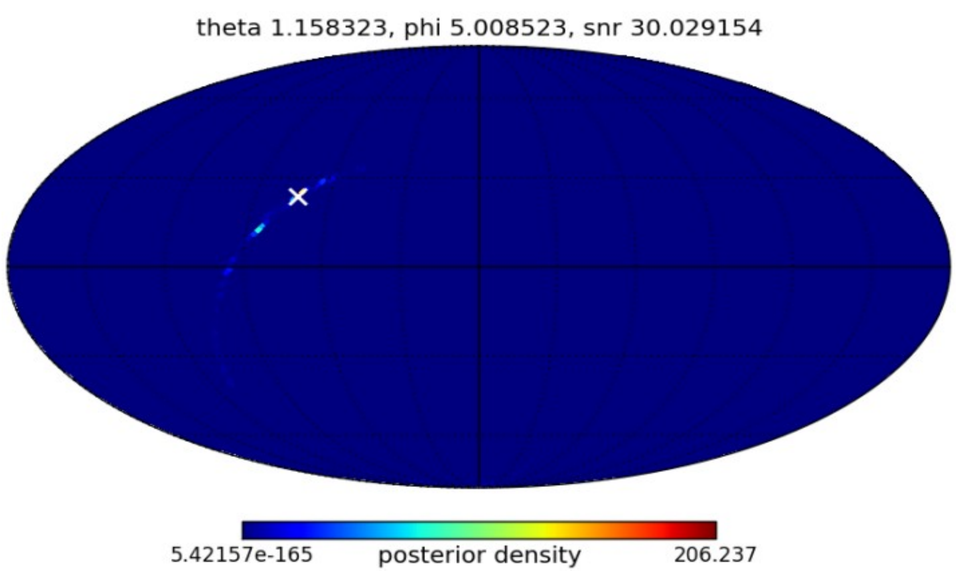
Preliminary Results ("Rectangular")

- Loud events

2-detector (HL) network



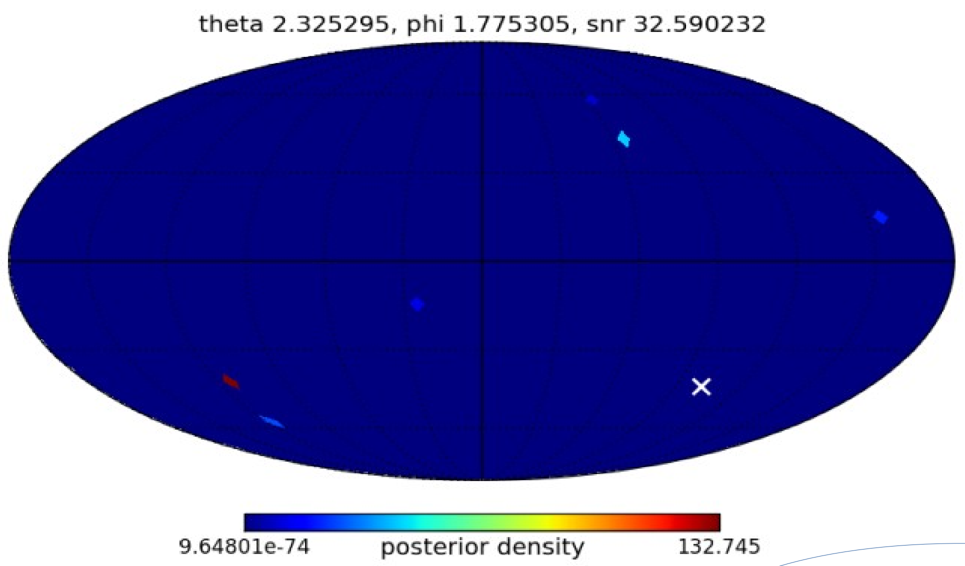
3-detector (HLV) network



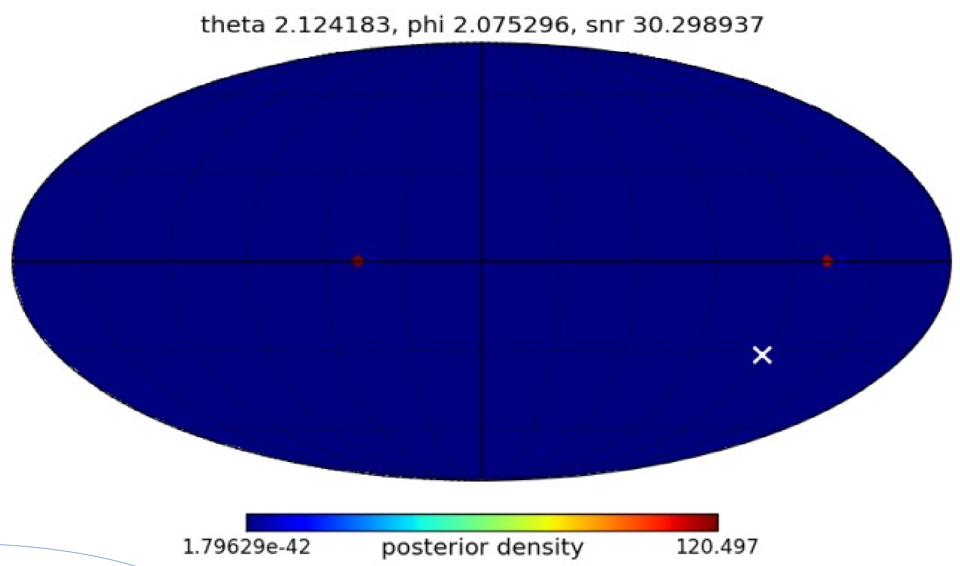
Preliminary Results ("Unmodeled")

- Threshold events:

2-detector (HL) network



3-detector (HLV) network



Overfitted data!

Conclusions and Future Outlook

- LIB
  - Already proven as parameter estimation tool (arXiv:1409.2435)
  - Detection pipeline using Bayes factor cuts looks promising, statistical study in the works
  - Should understand trade-offs with number of live points (latency, accuracy of evidence, accuracy of posterior)
- Low-latency pipeline
  - Preliminary results appear to be consistent with LIB with latencies of ~ (30 minutes) / (# CPUs)
  - Can implement better priors, strictly speaking $p(h_{rss}) \neq p(\{h_i\})$
  - Need to perform statistical tests to optimize model selection and compare to LIB results