Gravitational Wave Parameter Estimation

Neil J. Cornish

eXtreme Gravity Institute

MONTANA STATE UNIVERSITY

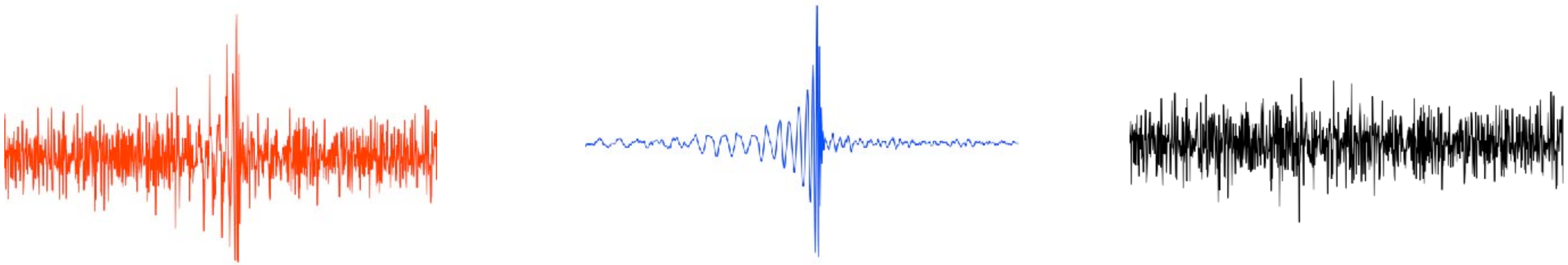
LSC LIGO Scientific Collaboration

VIRGO

Outline

- Signal analysis 101
- Bayesian Inference
- Bayesian Hierachical Modeling
- Noise modeling
- Tools of the trade
- Examples from LIGO/Virgo

Signals in additive noise

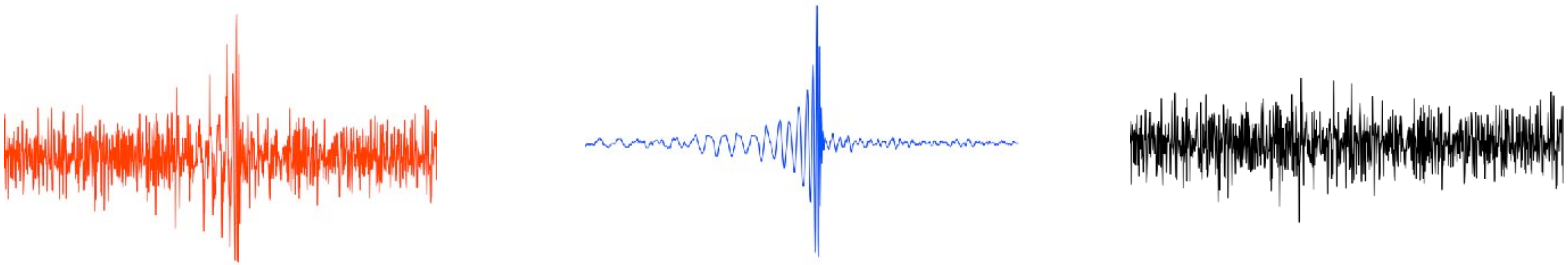


data = signal + noise

$$d = h + n$$

(randomly selected aLIGO data from 14 September 2015)

Its all about the residuals



data - signal = noise

$$d - h = n$$

$$p(d|h) = p(d - h) = p(n)$$

Noise models matter

$$p(d|h) = p(d - h) = p(n)$$

The likelihood is our statistical model for the noise

An incorrect noise model will bias the analysis.
e.g. assuming that noise is stationary and Gaussian when it is not.

Noise Models

$$p(d|h) = p(d-h) = p(n)$$

Example: Stationary, colored, Gaussian noise

$$p(n, S_n(f)) = \prod_f \frac{1}{2\pi S_n(f)} e^{-\frac{\tilde{n}_f \tilde{n}_f^*}{S_n(f)}}$$

Note: This is a parametrized model, depends on the unknown spectrum $S_n(f)$

Bayesian Inference

Noise model

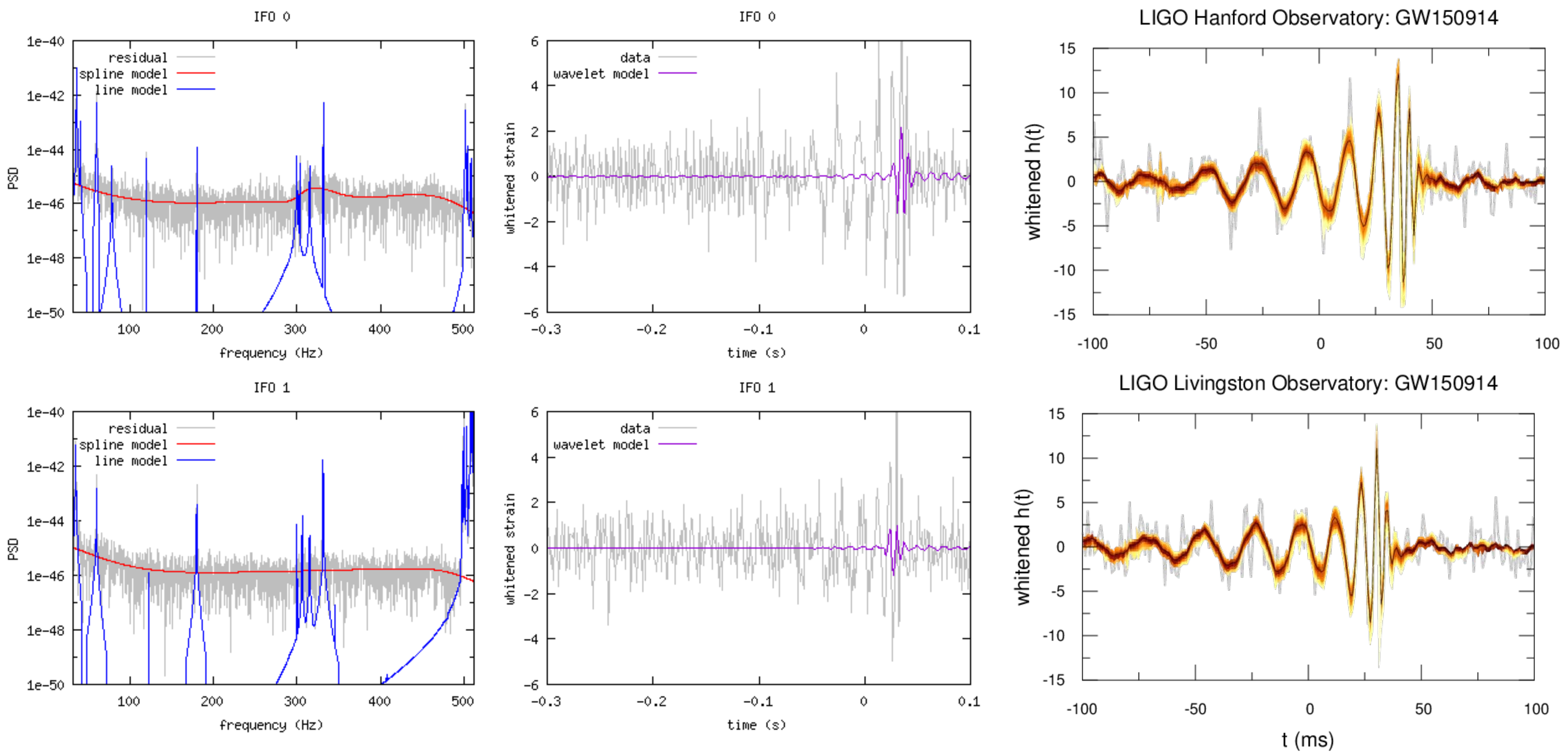
Signal model

Posterior probability for waveform h →

$$p(\mathbf{h}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{h})p(\mathbf{h})}{p(\mathbf{d})}$$

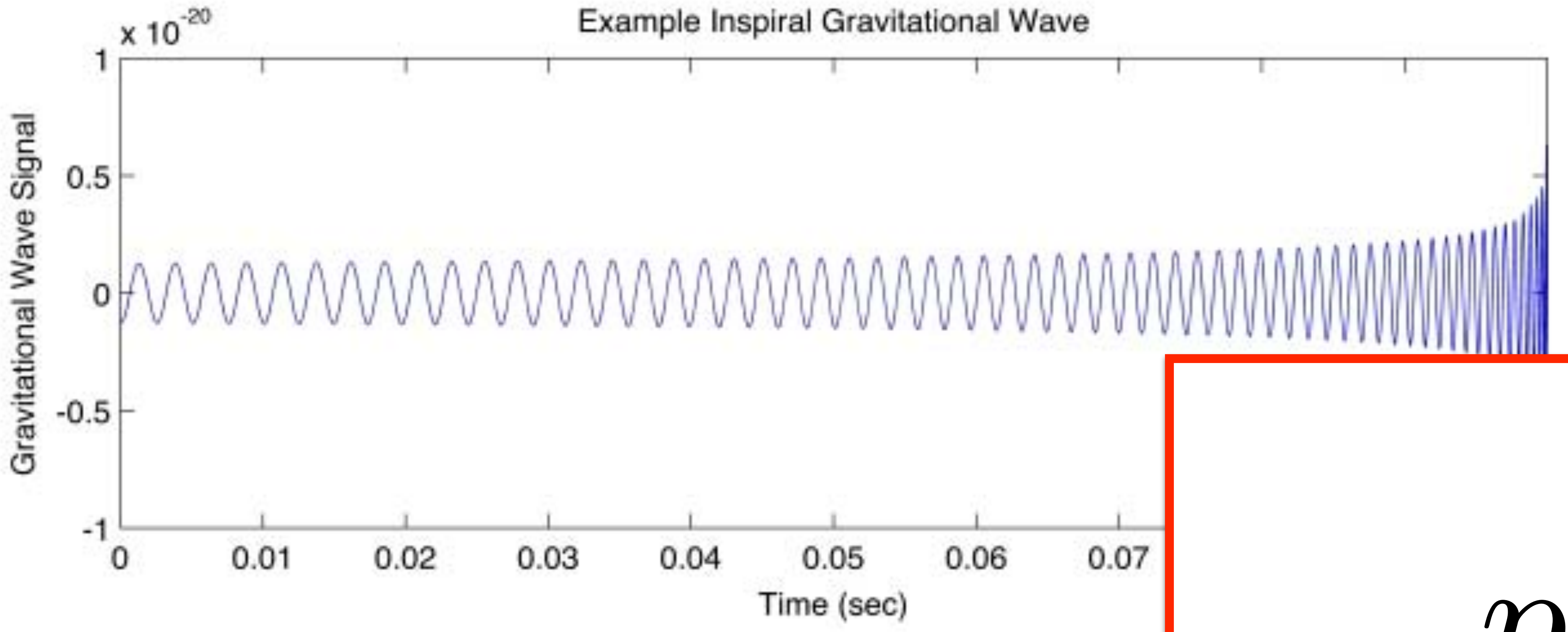
Normalization - model evidence

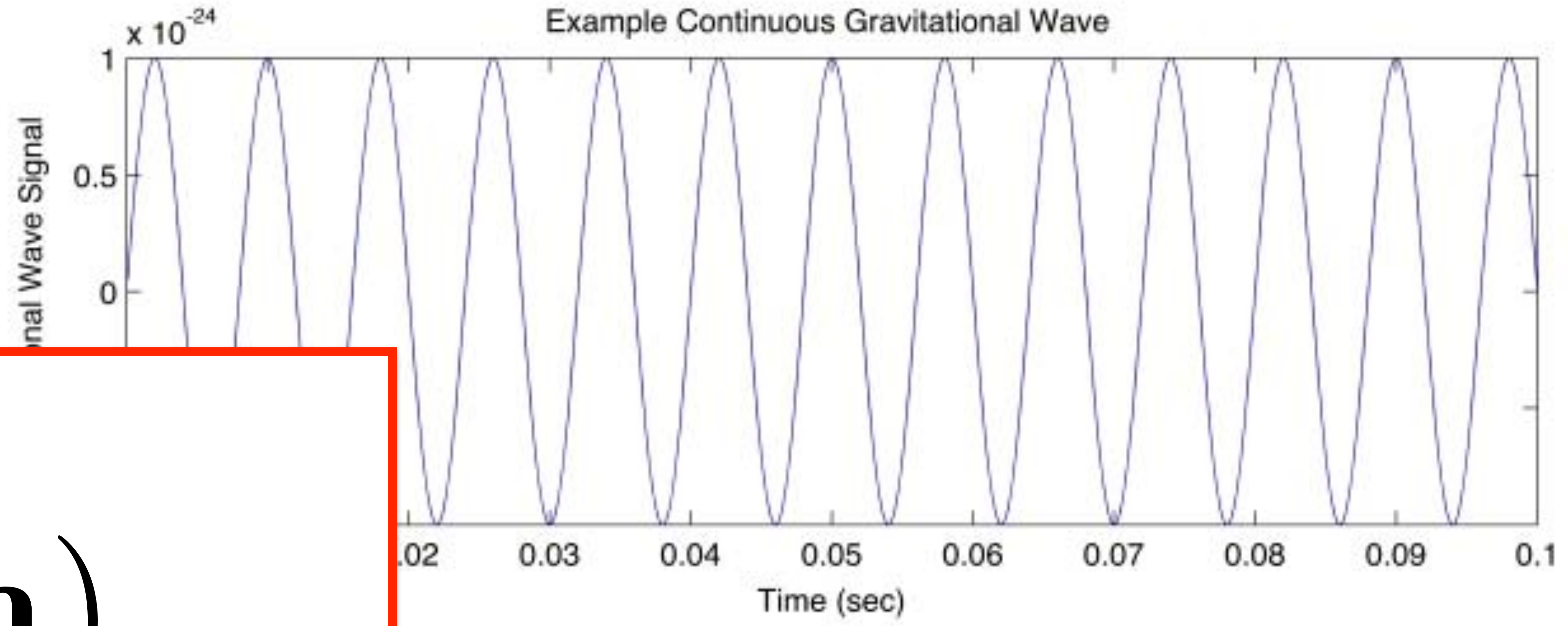
Reconstructing GW150914 with wavelets



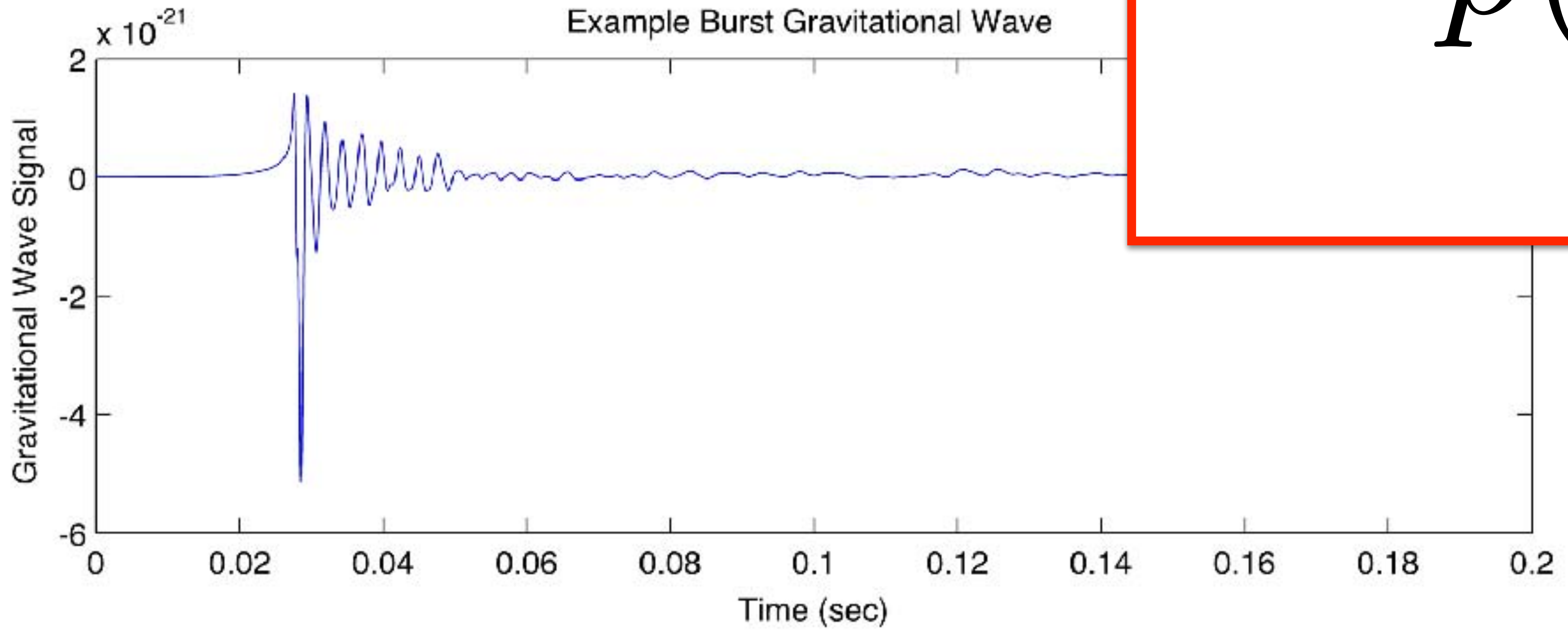
Gravitational wave signal types

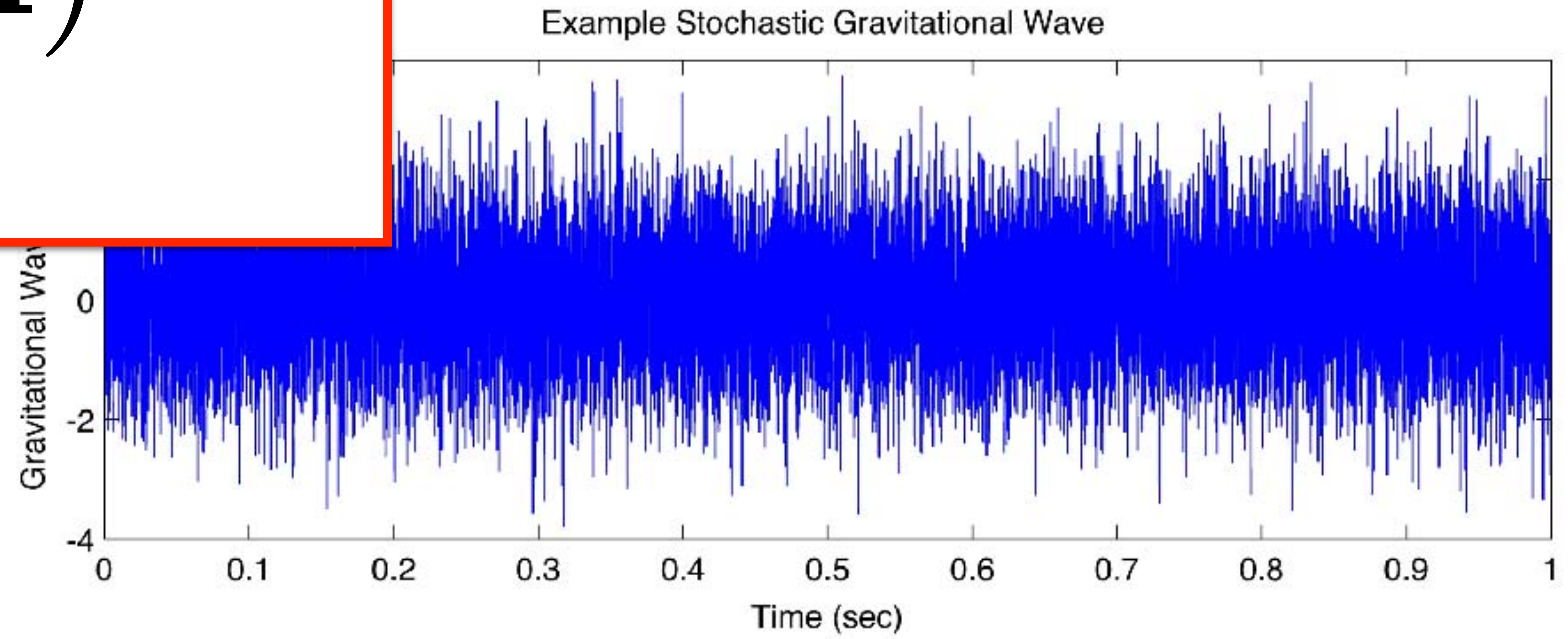
Well modeled - e.g. binary inspiral and merger





$$p(\mathbf{h})$$





Poorly modeled - e.g. core collapse supernovae

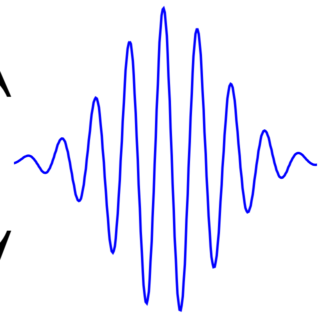
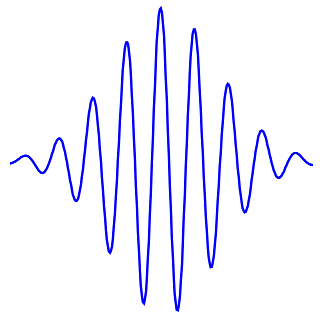
Stochastic- e.g. phase transition in early universe

Gravitational wave signal models

Template based

$$p(\mathbf{h}) = \delta(\mathbf{h} - \mathbf{h}(\vec{\lambda})), \quad p(\vec{\lambda})$$

Burst signals

$$p(\mathbf{h}) = \delta\Big(\mathbf{h} - \sum$$

$$\Big), \quad p\Big($$

$$\Big)$$

Stochastic signals

$$p(\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi \mathbf{S_h})}} e^{-\frac{1}{2}(\mathbf{h}^\dagger \mathbf{S_h}^{-1} \mathbf{h})}, \quad p(\mathbf{S_h})$$

Bayesian Hierachical Modeling

Level 1 Inference

$$p(\mathbf{h}|\mathbf{d},\mathcal{H}_i,\boldsymbol{\theta}) = \frac{p(\mathbf{d}|\mathbf{h},\mathcal{H}_i,\boldsymbol{\theta})p(\mathbf{h}|\mathcal{H}_i,\boldsymbol{\theta})}{p(\mathbf{d}|\mathcal{H}_i,\boldsymbol{\theta})}$$

We have some model $\mathcal{H}_i$ for the signal and the noise described by the parameters $\boldsymbol{\theta}$

The result of the analysis is a posterior distribution for the waveform, in addition to posterior distributions for the sky location, polarization and the noise properties

BayesWave performs Level 1 Inference

Bayesian Hierachical Modeling

But what if we are not really interested in the waveform samples $\mathbf{h}$?

Marginalize over $\mathbf{h}$

$$p(\mathbf{d}|\mathcal{H}_i, \boldsymbol{\theta}) = \int p(\mathbf{d}|\mathbf{h}, \mathcal{H}_i, \boldsymbol{\theta}) p(\mathbf{h}|\mathcal{H}_i, \boldsymbol{\theta}) d\mathbf{h}$$

The marginal likelihood from Level 1 becomes the likelihood for Level 2

Level 2 Inference

$$p(\boldsymbol{\theta}|\mathbf{d}, \mathcal{H}_i) = \frac{p(\mathbf{d}|\mathcal{H}_i, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{H}_i)}{p(\mathbf{d}|\mathcal{H}_i)}$$

LALinference performs Level 2 Inference (Can recover Level 1 in post-processing)

Bayesian Hierachical Modeling

But what if we are not all that interested in the model parameters $\boldsymbol{\theta}$?

Marginalize over $\boldsymbol{\theta}$

$$p(\mathbf{d}|\mathcal{H}_i) = \int p(\mathbf{d}|\mathcal{H}_i, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{H}_i) d\boldsymbol{\theta}$$

The marginal likelihood from Level 2 is the model evidence. It can be used for model selection

Level 3 Inference

$$O_{ij} = \frac{p(\mathcal{H}_i|\mathbf{d})}{p(\mathcal{H}_j|\mathbf{d})} = \frac{p(\mathcal{H}_i)}{p(\mathcal{H}_j)} \frac{p(\mathbf{d}|\mathcal{H}_i)}{p(\mathbf{d}|\mathcal{H}_j)}$$

Prior Odds ↑ Bayes Factor ↑

Level 2 Inference: Template based analysis

These have the strongest priors and hence yield the most sensitive searches

$$p(\mathbf{h}) = \delta(\mathbf{h} - \mathbf{h}(\vec{\lambda})), \quad p(\vec{\lambda})$$

Integrating out the individual signal samples yields the marginal likelihood in terms of the model parameters

$$p(\mathbf{d}|\vec{\lambda}) = \int p(\mathbf{d}|\mathbf{h})\delta(\mathbf{h} - \mathbf{h}(\vec{\lambda}))\, d\mathbf{h}$$

The marginal likelihood and the (hyper) prior on the model parameters then defines the posterior on the model parameters

Likelihood **Prior**

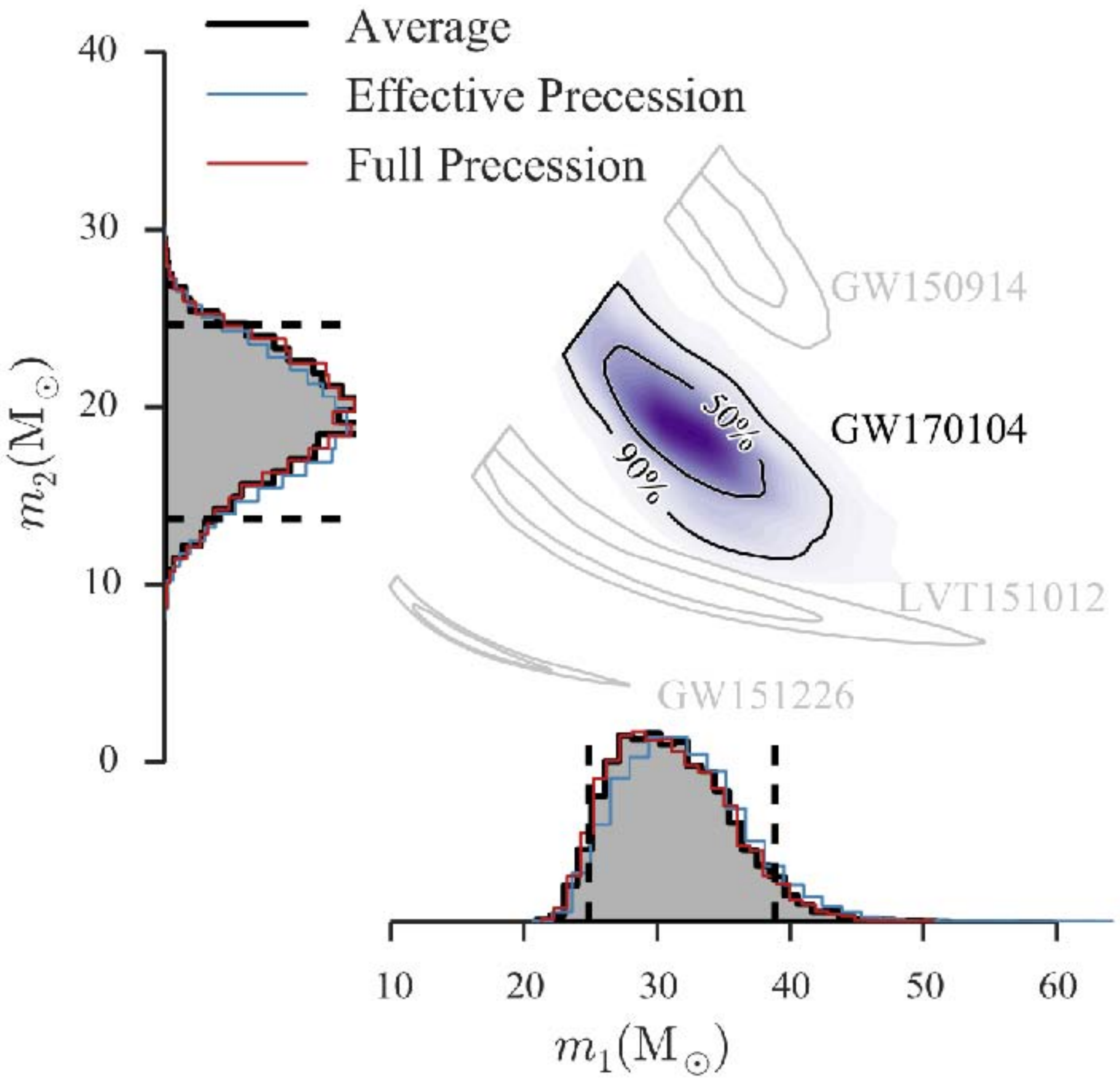
$$p(\vec{\lambda}|\mathbf{d}) = \frac{p(\mathbf{d}|\vec{\lambda})p(\vec{\lambda})}{p(\mathbf{d})}$$

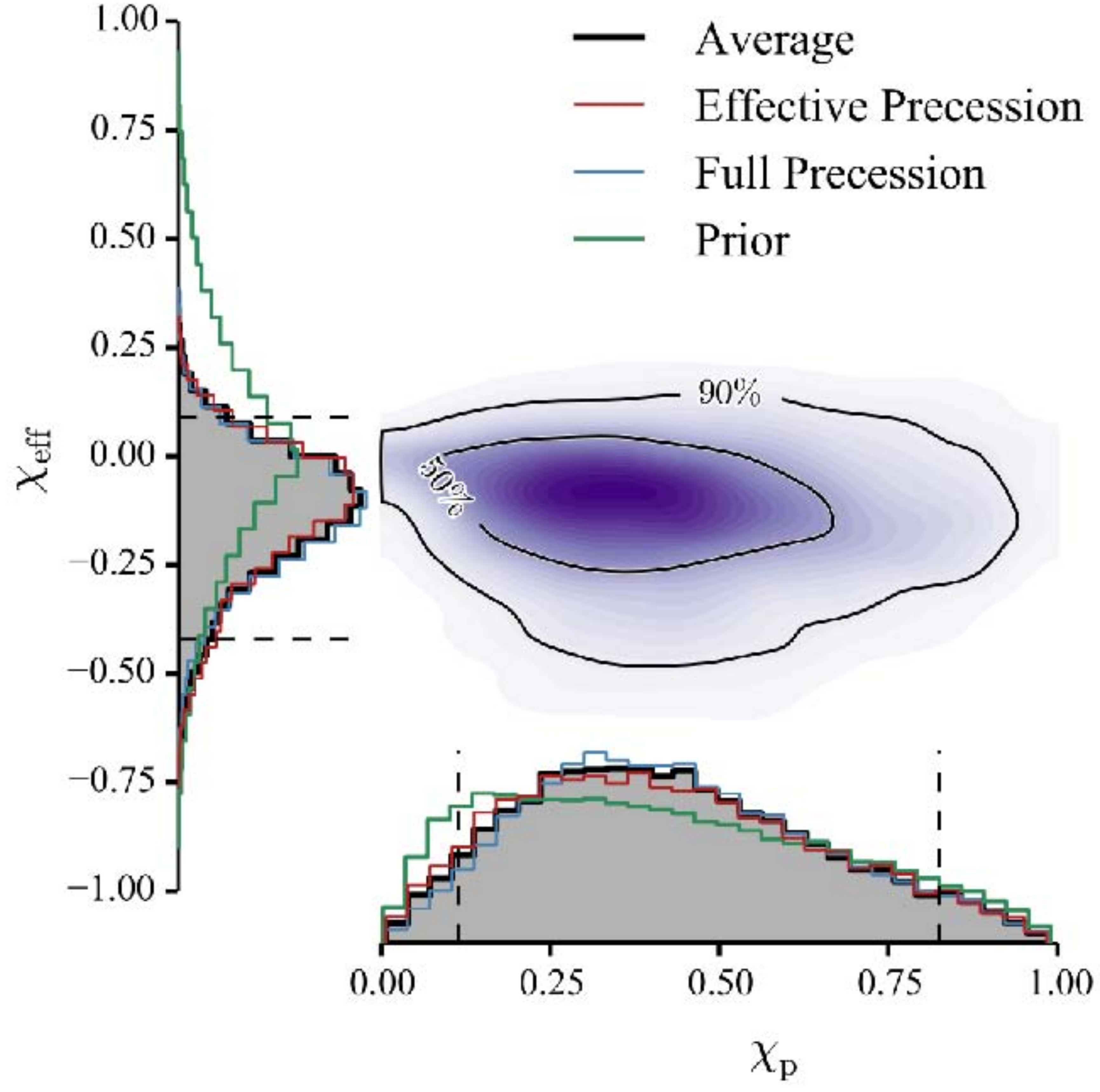
Evidence

Techniques such as MCMC and Nested Sampling can be used to map out the full posterior distribution, allowing us to compute mean, median and mode and credible intervals.

Template based analysis - Parameter Posteriors

$$p(\vec{\lambda}|\mathbf{d})$$





Level 2 Inference: Stochastic Signals

Gaussian random waveform

$$p(\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi \mathbf{S_h})}} e^{-\frac{1}{2}(\mathbf{h}^\dagger \mathbf{S_h}^{-1} \mathbf{h})}, \quad p(\mathbf{S_h})$$

Gaussian likelihood. The noise correlation matrix block diagonal between detectors

$C_{(Ik)(Jl)} = \delta_{IJ} (\mathbf{S_n})_{kl}$

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi \mathbf{C})}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{h})^\dagger \mathbf{C}^{-1} (\mathbf{d}-\mathbf{h})}$$

Marginal likelihood

$$p(\mathbf{d}|\mathbf{S_h}) = \int p(\mathbf{d}|\mathbf{h}) p(\mathbf{h})\, d\mathbf{h} = \frac{1}{\sqrt{\det(2\pi \mathbf{G})}} e^{-\frac{1}{2}(\mathbf{d}^\dagger \mathbf{G}^{-1} \mathbf{d})}$$

$$G_{(Ik)(Jl)} = \delta_{IJ} (\mathbf{S_n})_{kl} + \kappa_{(Ik)(Jl)} (\mathbf{S_h})$$

Overlap reduction function, Hellings-Downs curve

[Cornish & Romano, PRD **87** 122003 (2013)]

Level 3 Inference: Rate of Black Hole Mergers

Likelihood for duty cycle

$$p(\mathbf{d}|\xi) = \prod_{j=1}^{M} \left[\xi\, p(\mathbf{d}_j|\mathrm{Signal}) + (1-\xi)\, p(\mathbf{d}_j|\mathrm{Noise})\right]$$

Marginal Likelihood (Evidence) for Signal Model in data chunk j

Marginal Likelihood (Evidence) for Noise Model in data chunk j

Fraction of chunks with mergers ξ

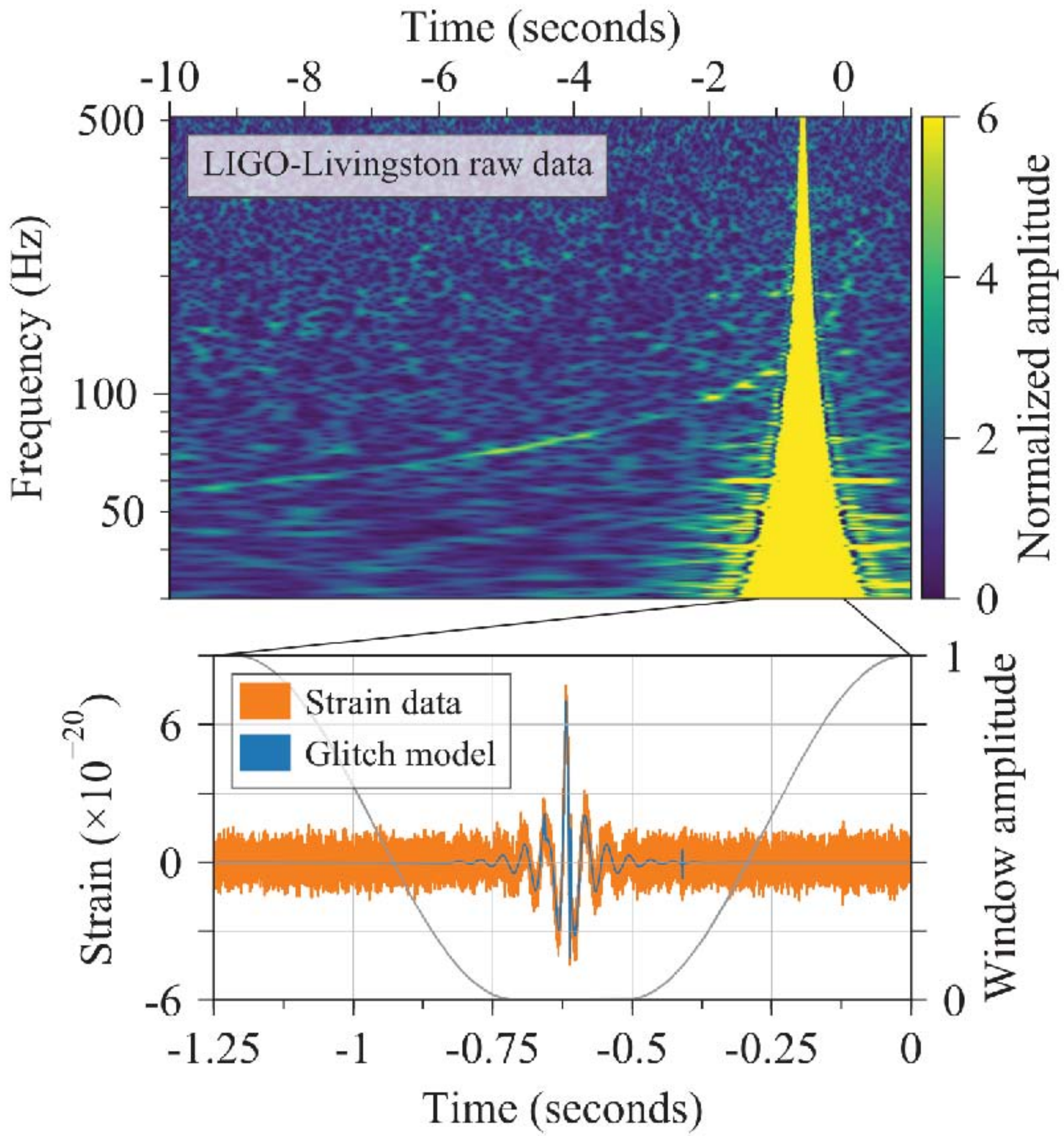
Number of binary mergers $N = M\xi$

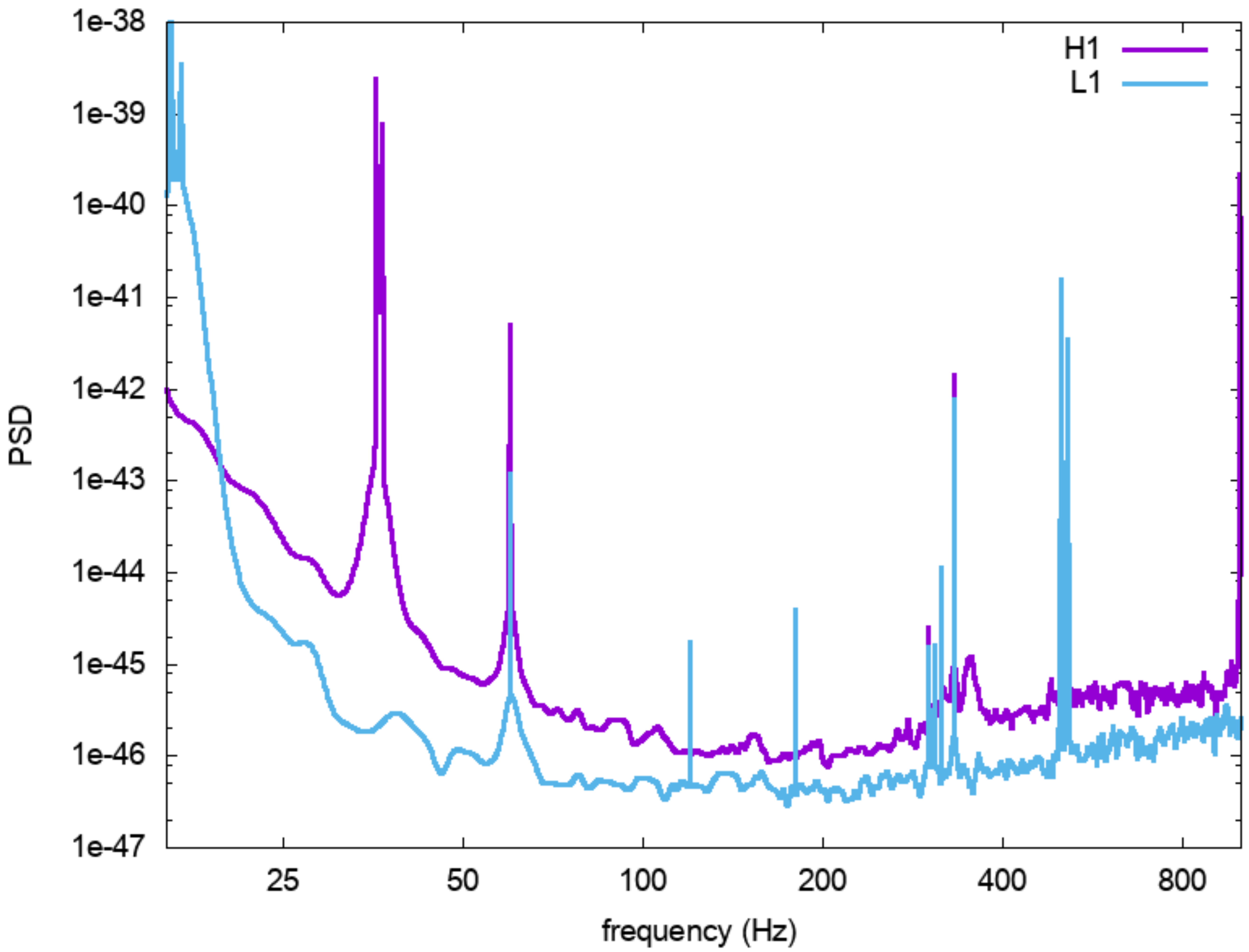
Likelihood for merger rate

$$p(\mathbf{d}|R) = M \sum_{N} \int d\xi p\left(\mathbf{d}|\xi\right) p(N) \frac{e^{-R} R^N}{N!}$$

[Smith & Thrane, PRX **kk** (2018); Cornish, Physics (2018)]

Noise Models





Spectra computed using 8 second segments, spaced 2 seconds apart, covering 256 surrounding the BNS merger (de-glitched data)

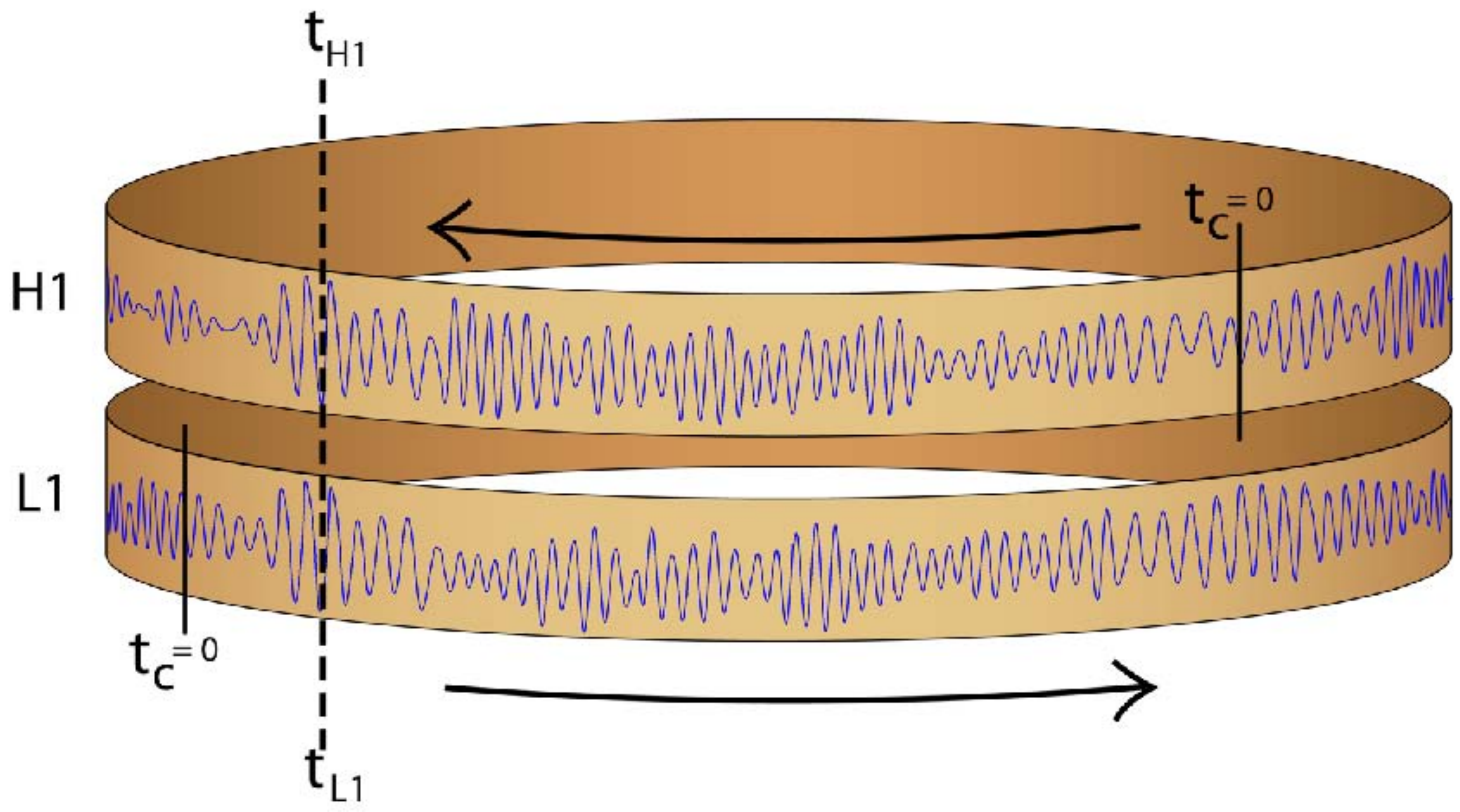
GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, PRL **119** 161101 (2017)

Contending with non-stationary, non-Gaussian noise: Traditional approach

Non-stationary: → Adiabatic drifts in the PSD — - work with short data segments

Non-stationary: → Glitches — - vetoes and time-slides

Non-Gaussian: → Glitches



Contending with non-stationary, non-Gaussian noise: Bayesian approach

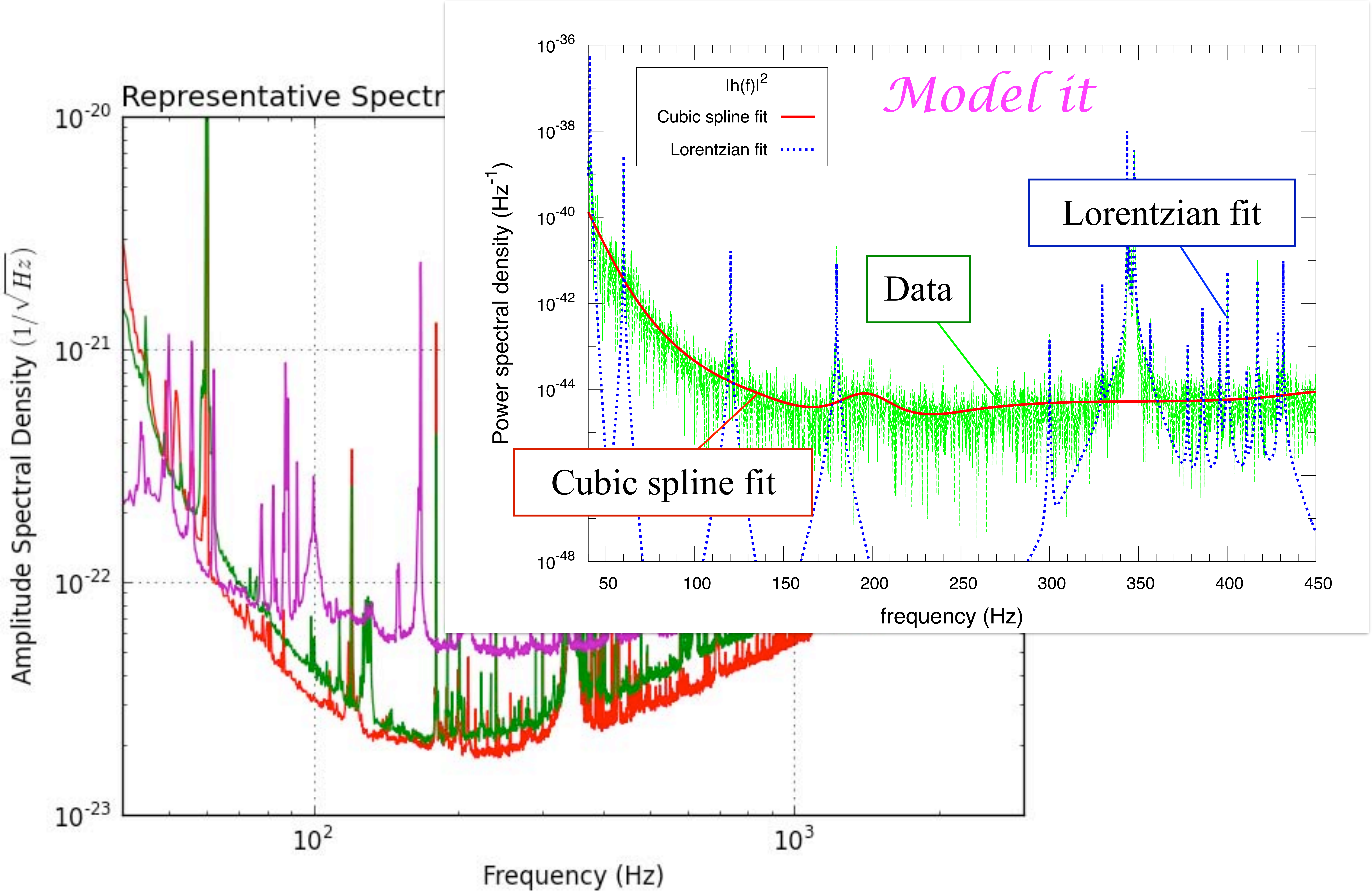
BayesWave

- Bayesian model selection
  - Three part model (signal, glitches, Gaussian noise)
  - Trans-dimensional Markov Chain Monte Carlo
- Wavelet decomposition
  - Glitches modeled by wavelets
  - Number, amplitude, duration and location of wavelets varies

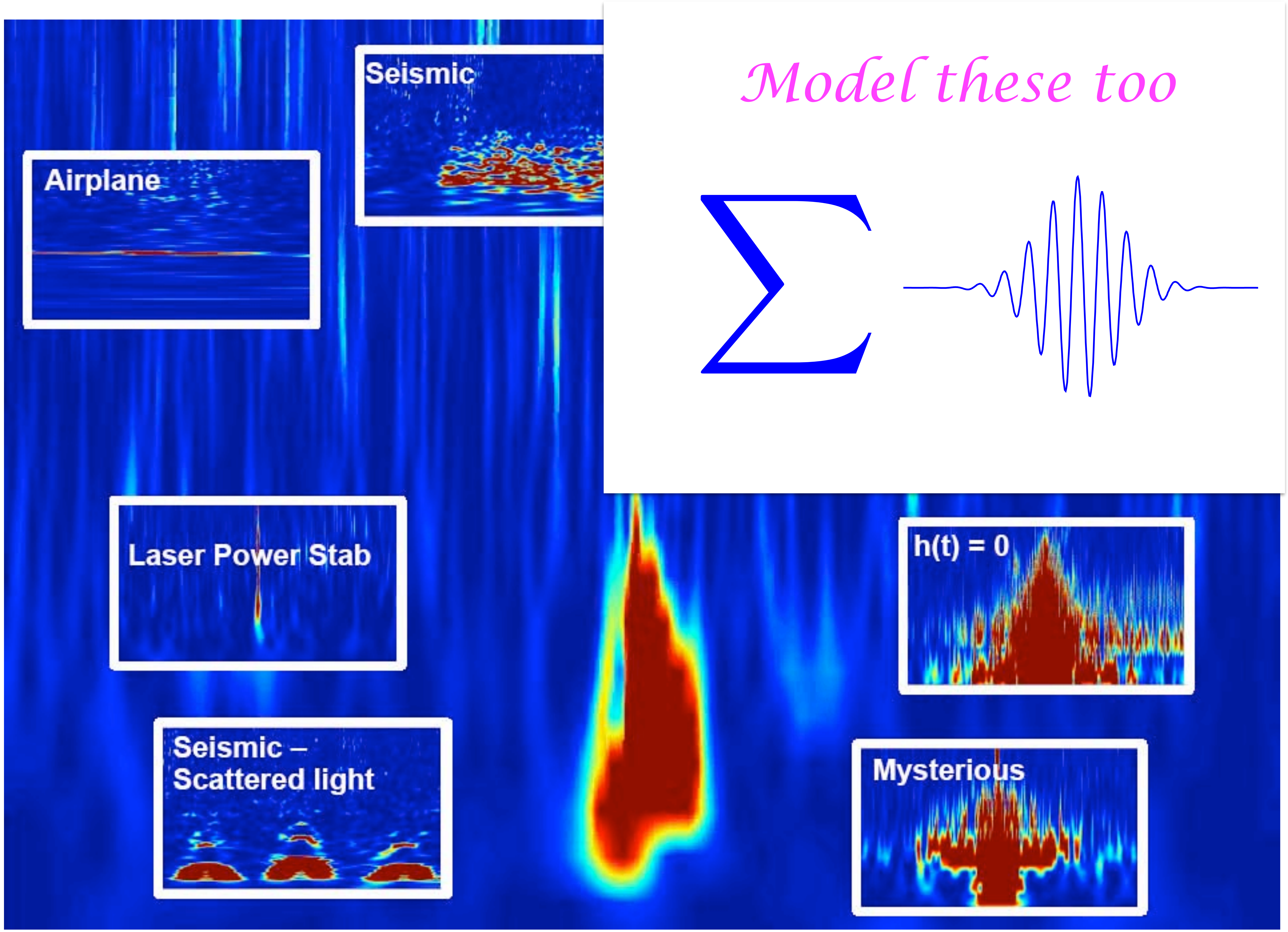
(Currently the model does not account for adiabatic drifts in the PSD)

[Cornish & Littenberg, Class. Quant. Grav. **32** 135012 (2015)]

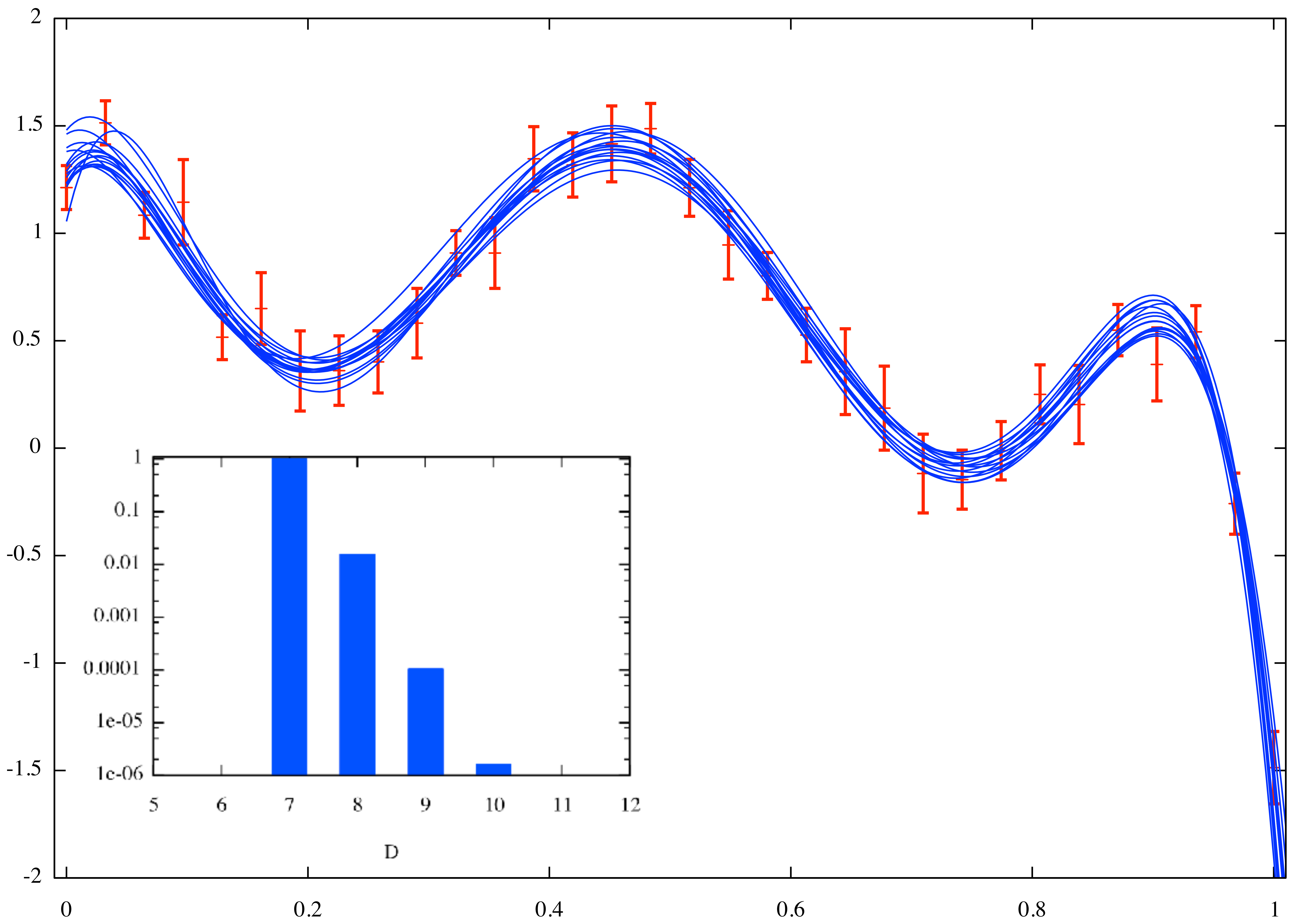
Spectral Modeling



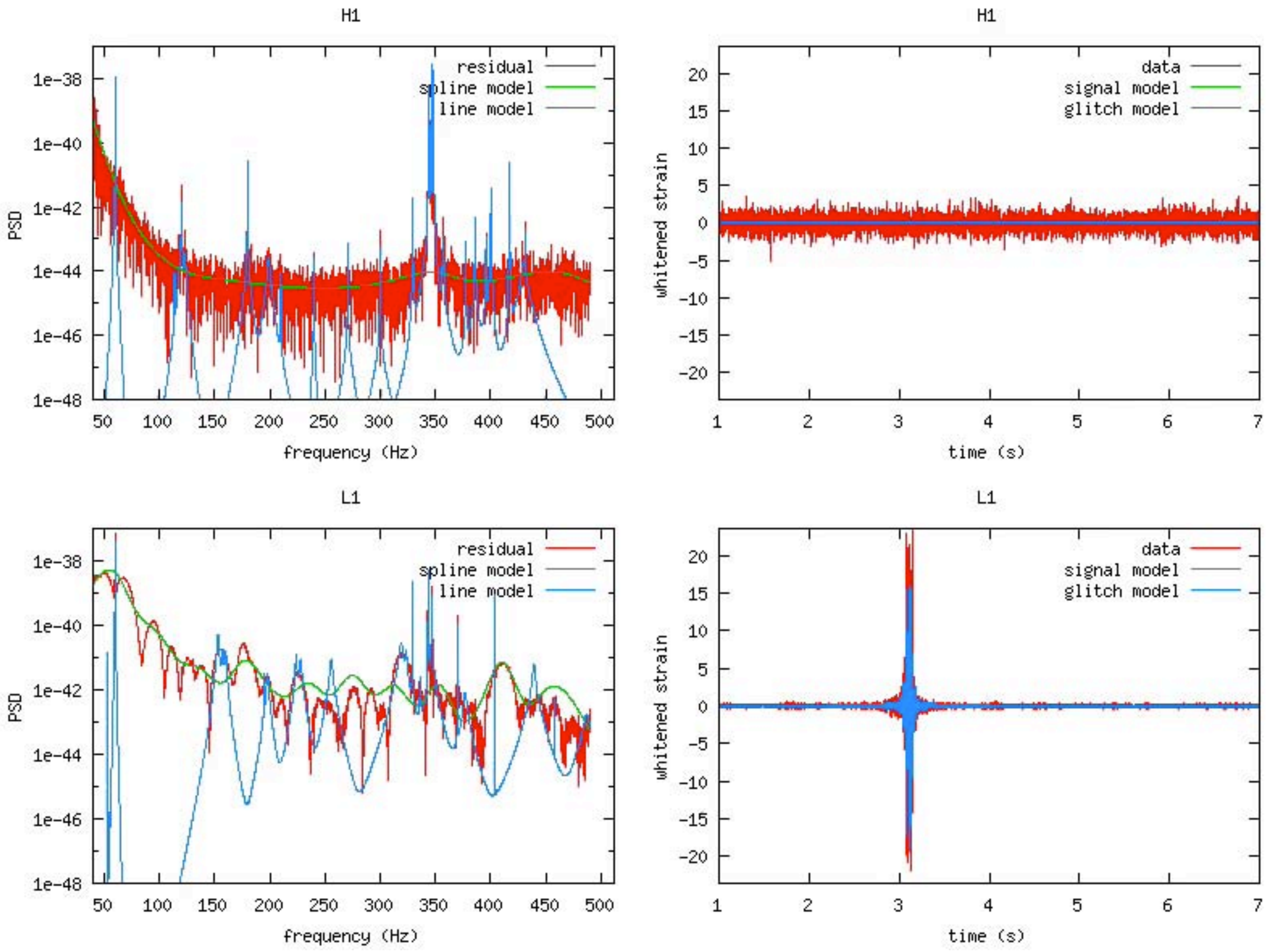
Glitches


Seismic
Airplane
Laser Power Stab
h(t) = 0
Seismic –
Scattered light
Mysterious
Model these too

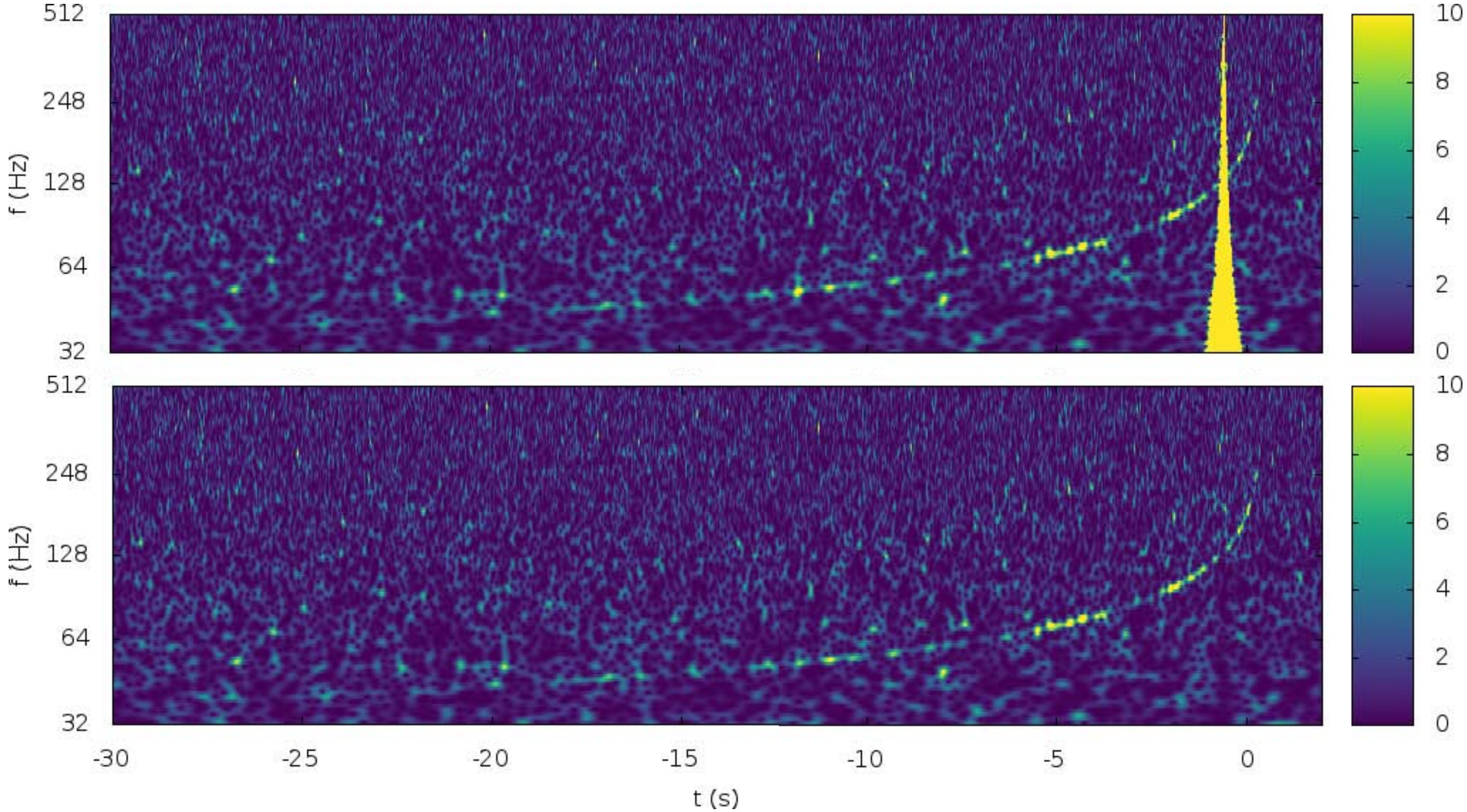
Trans-dimensional Markov Chain Monte Carlo



Example from LIGO S5 science run



Removing a glitch was preferable to vetoing GW170817



Adiabatic drifts in the PSD - Locally Stationary Noise

Likelihood for non-stationary Gaussian Noise

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{h})^\dagger \mathbf{C}^{-1}(\mathbf{d}-\mathbf{h})}$$

Cost of computing the likelihood is far less in a representation where the noise correlation matrix **C** is diagonal

e.g. Colored stationary noise has a diagonal noise correlation matrix in the Fourier domain

Pulsar Timing has to deal with colored, non-stationary data and un-even sampling - analysis performed directly in the time domain. Clever tricks have been developed to speed up the costly matrix inversions and sums

Likelihood for Non-stationary Gaussian Noise

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{h})^\dagger \mathbf{C}^{-1}(\mathbf{d}-\mathbf{h})}$$

Cost of computing the likelihood is far less in a representation where the noise correlation matrix **C** is diagonal

For a large class of discrete wavelet transformations and locally stationary noise [1]

$$C_{(i,j)(k,l)} = \delta_{ij}\delta_{kl}\, C_{ik}$$ [2]

Time Frequency

This is the likelihood used by the LIGO coherent WaveBurst algorithm

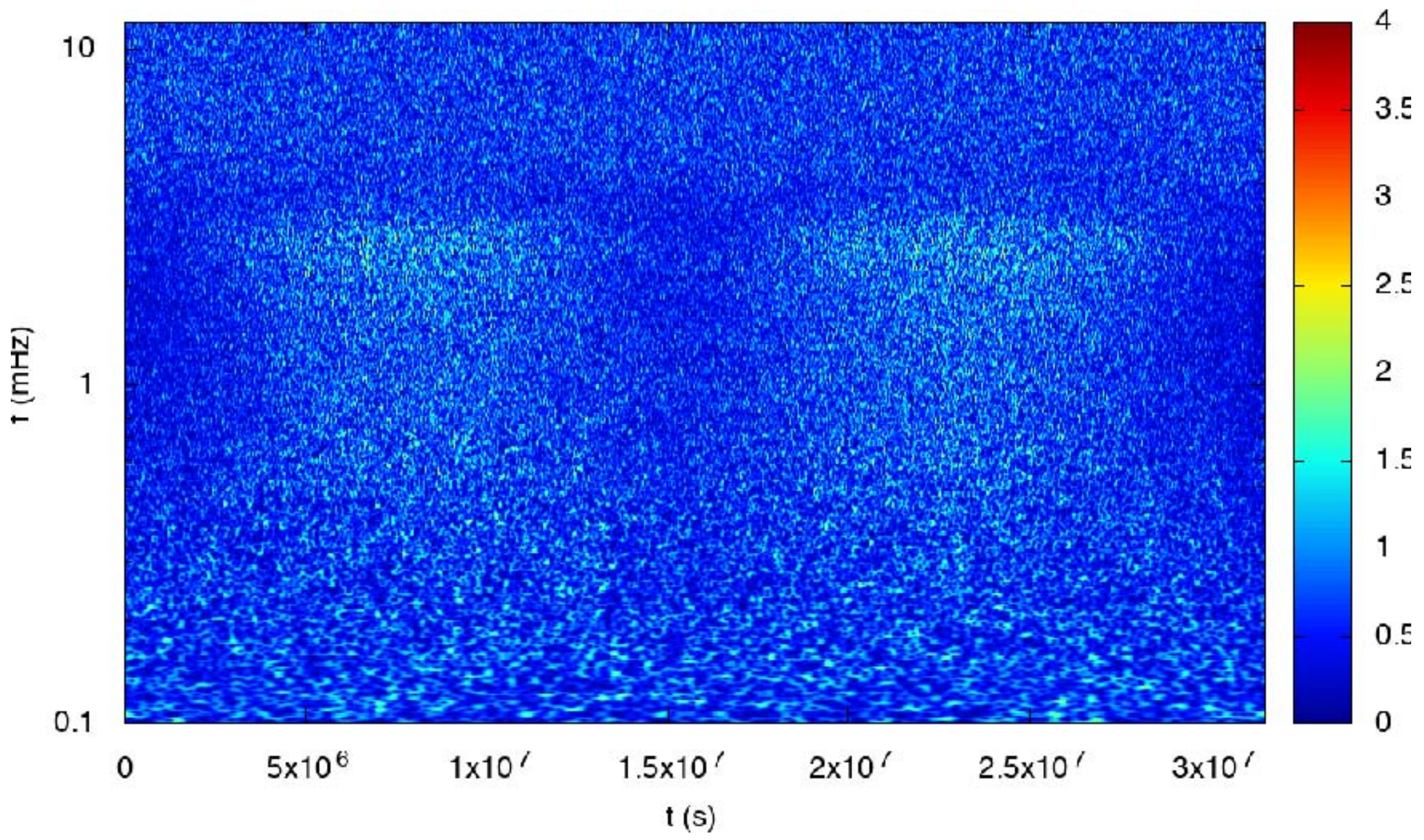
[**1.** "Fitting time series models to nonstationary processes". Dahlhaus, Ann. Statist., 25, 1 (1997)]

[**2.** "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum", Nason, von Sachs, & Kroisandt, J. R. Statist. Soc. Series B**62**, 271 (2000)]

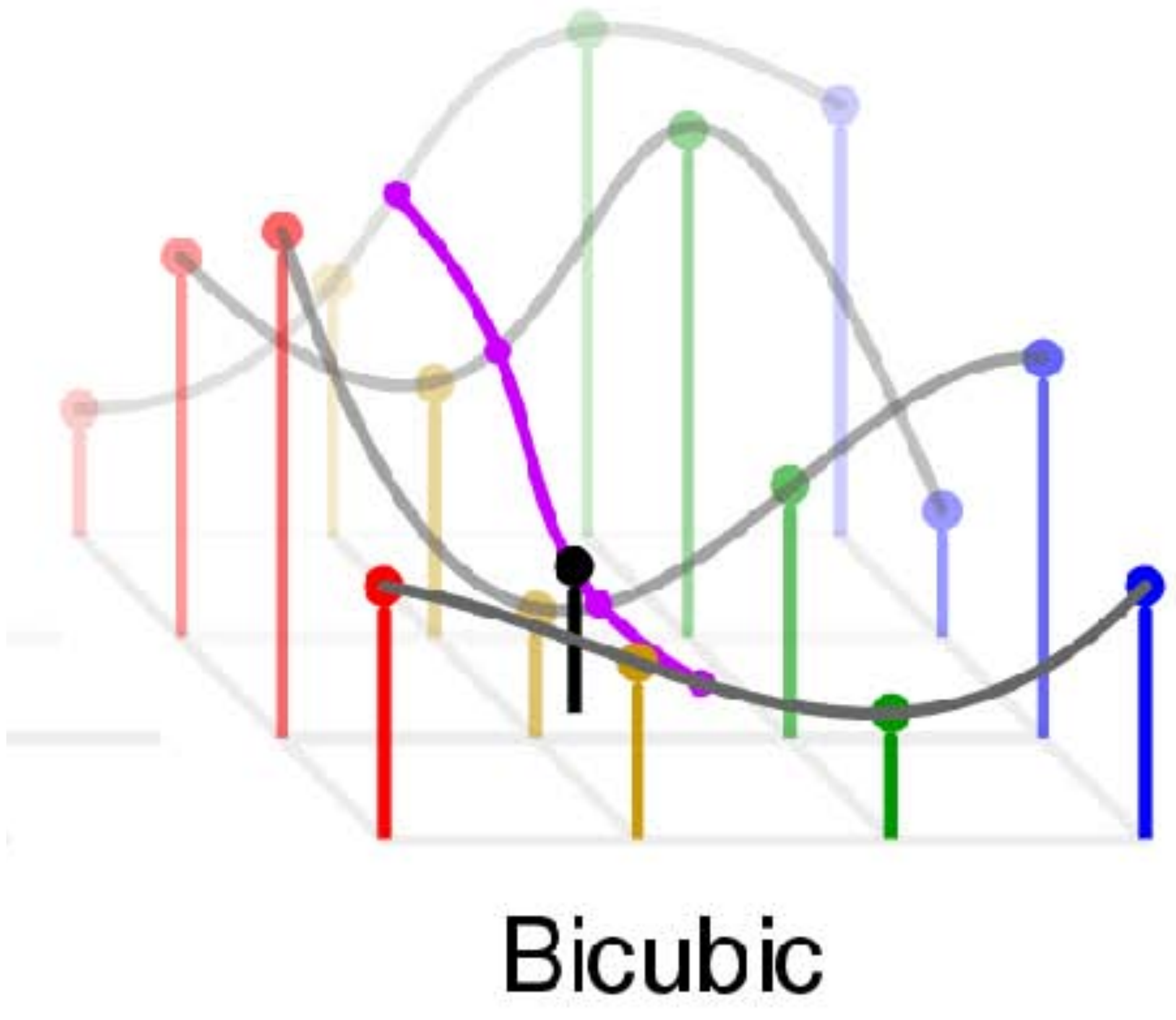
Likelihood for Non-stationary Gaussian Noise

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{h})^{\dagger}\mathbf{C}^{-1}(\mathbf{d}-\mathbf{h})}$$

Can we use a discrete wavelet based likelihood? $C_{(i,j)(k,l)} = \delta_{ij}\delta_{kl}\, C_{ik}$



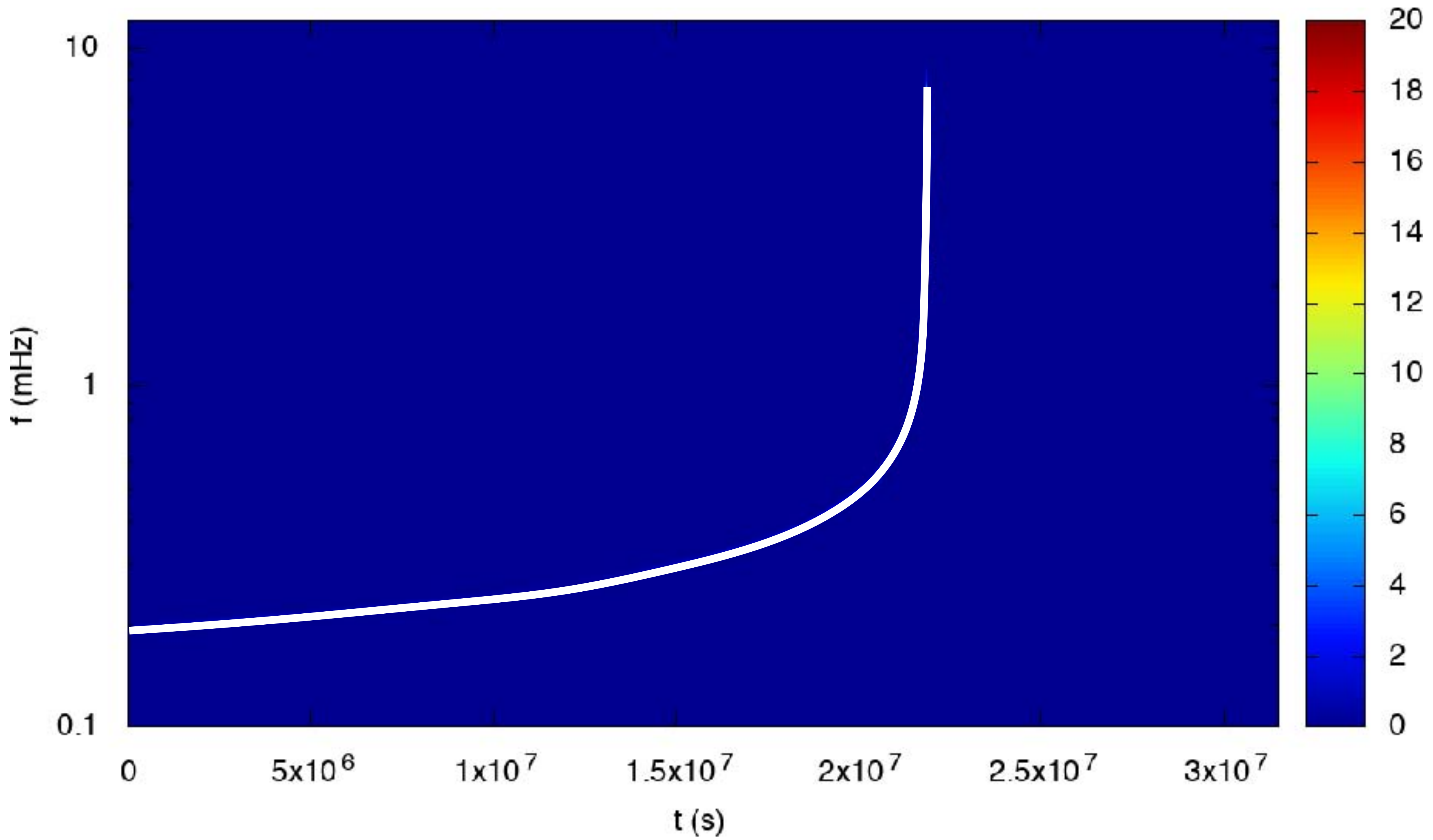
Model the wavelet spectrum C_{ik} as a smooth function in frequency and time. E.g. Trans-dimensional Bicubic spline



Likelihood for Non-stationary Gaussian Noise

$$p(\mathbf{d}|\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{h})^\dagger \mathbf{C}^{-1} (\mathbf{d}-\mathbf{h})}$$

Can we use a discrete wavelet based likelihood? $C_{(i,j)(k,l)} = \delta_{ij}\delta_{kl} C_{ik}$



Need fast wavelet transforms of the signals for computational efficiency

Use SPA to derive analytic wavelet domain waveforms?

Reduce order modeling?

Only compute wavelets along predicted t-f track?

Bayesian Inference: Tools of the Trade

Prior

$p(h|M)$

Likelihood

$p(d|h)$

$p(h|d, M)$

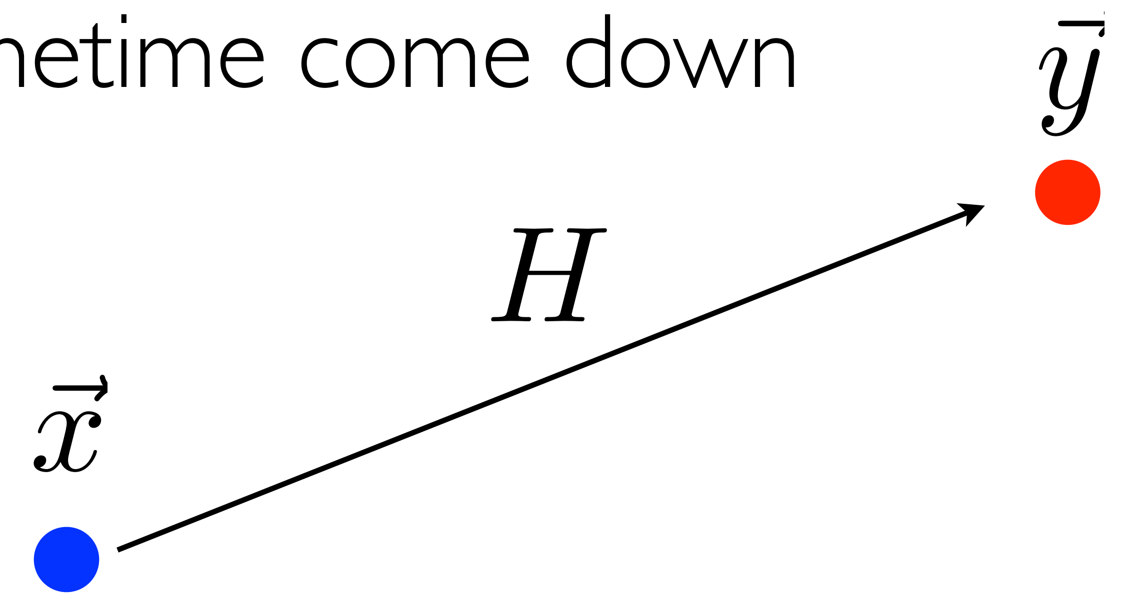
Posterior

$p(d|M)$

Evidence

Markov Chain Monte Carlo

Always go up,
Sometime come down



Yields PDF $p(\vec{x}|d)$ for parameters $\vec{x}$ given data d

$$H = \min\left(1, \frac{p(\vec{y})\,p(d|\vec{y})\,q(\vec{x}|\vec{y})}{p(\vec{x})\,p(d|\vec{x})\,q(\vec{y}|\vec{x})}\right)$$

Prior — Likelihood — Proposal

Transition Probability
(Metropolis-Hastings)

MCMC Recipe

Ingredients:

Local posterior approximation

Global likelihood maps

Differential evolution proposals

Parallel tempering

Directions:

Mix all the proposals together. Check consistency by recovering the prior and producing diagonal PP plots. Results are ready when distributions are stationary.

Proposal Distributions

Local posterior approximation

Quadratic approximation to the posterior using the augmented Fisher Information Matrix

$$q(\vec{y}|\vec{x}) = \frac{1}{\sqrt{\det(2\pi\mathbf{K}^{-1})}} e^{-\frac{1}{2}K_{ij}(x^i - y^i)(x^j - y^j)}$$

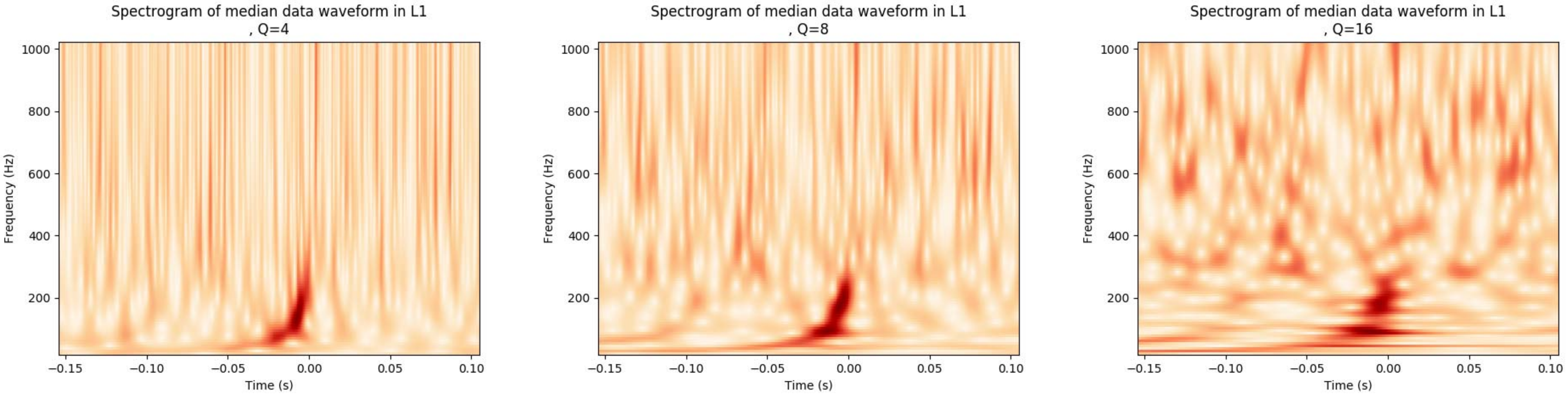
Propose jumps along eigendirections of **K**, scaled by eigenvalues

Global likelihood maps

Use a Non-Markovian Pilot search (hill climbers, simulated annealing, genetic algorithms etc) to crudely map the posterior/likelihood and use this as a proposal distribution for a Markovian follow-up [Littenberg & Cornish, PRD 80, 063007, (2009)]

Time-frequency maps, Maximized likelihood maps

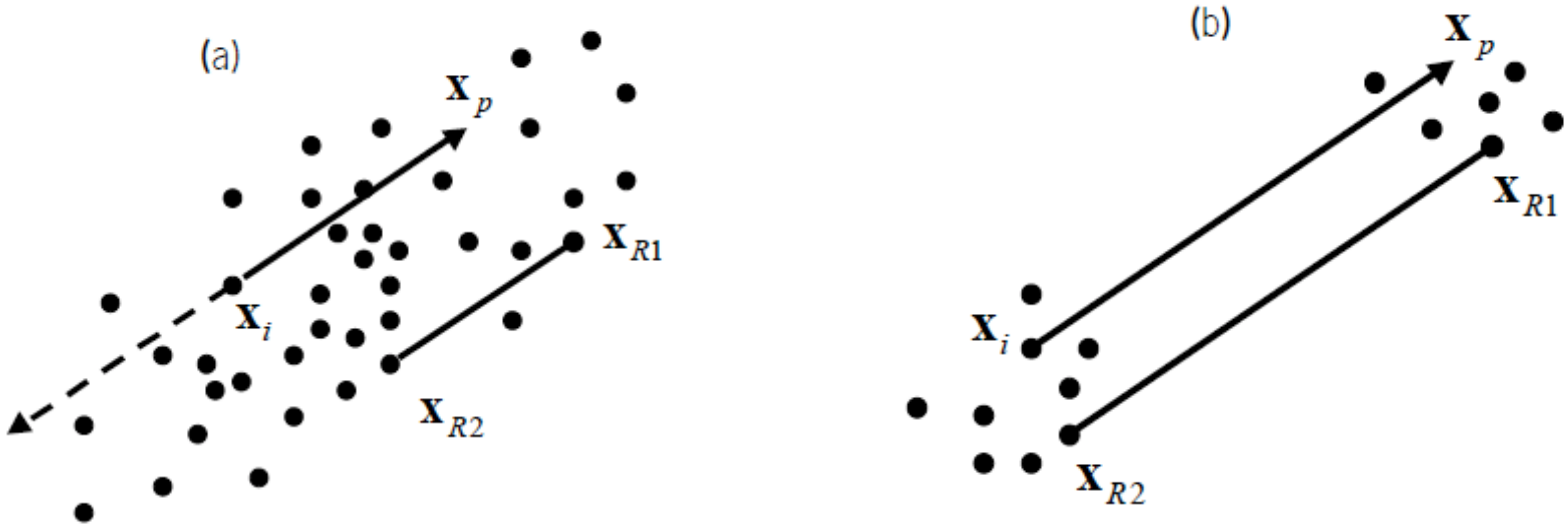
BayesWave Global Map Proposal



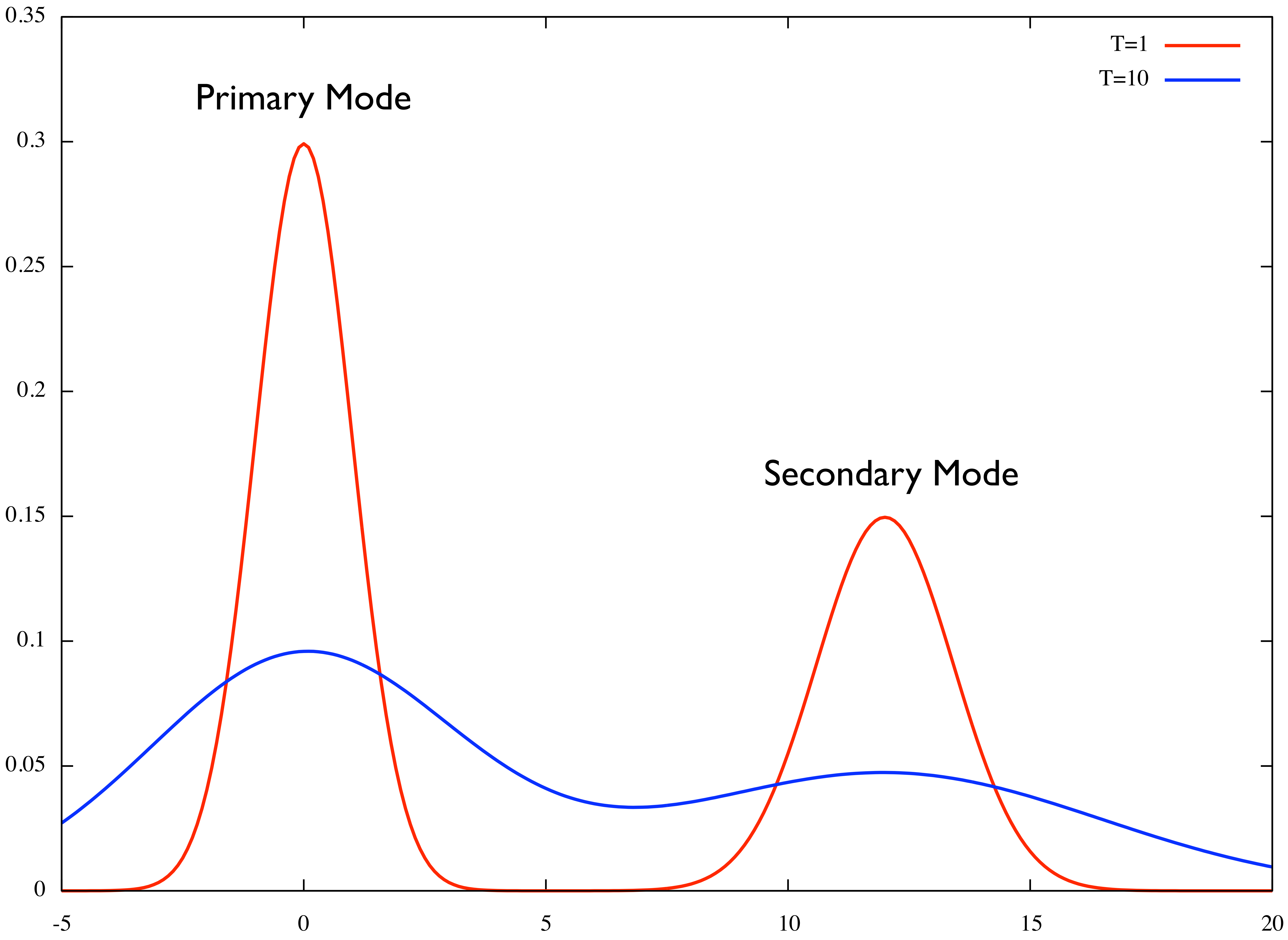
Proposal Distributions

Differential evolution

[Braak (2005)]



Parallel Tempering [Swendsen & Wang, 1986]



Ordinary MCMC techniques side-step the need to compute the evidence. PT uses multiple, coupled chains to improve mixing, and also allows the evidence to be computed.

Explore tempered posterior

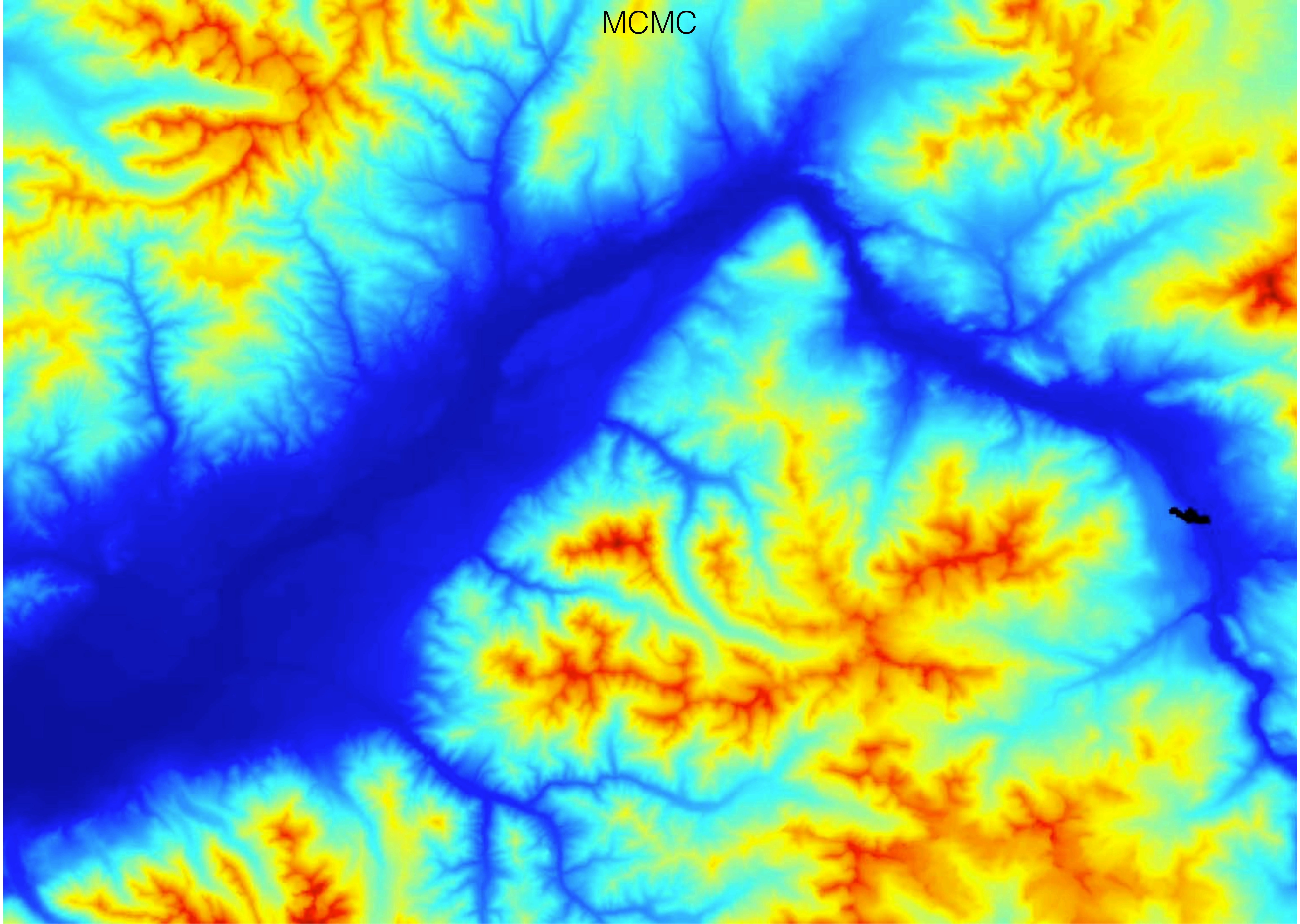
$$\pi(\vec{\lambda}|\mathbf{d})_T = p(\mathbf{d}|\vec{\lambda})^{1/T} p(\vec{\lambda})$$

Compute model evidence

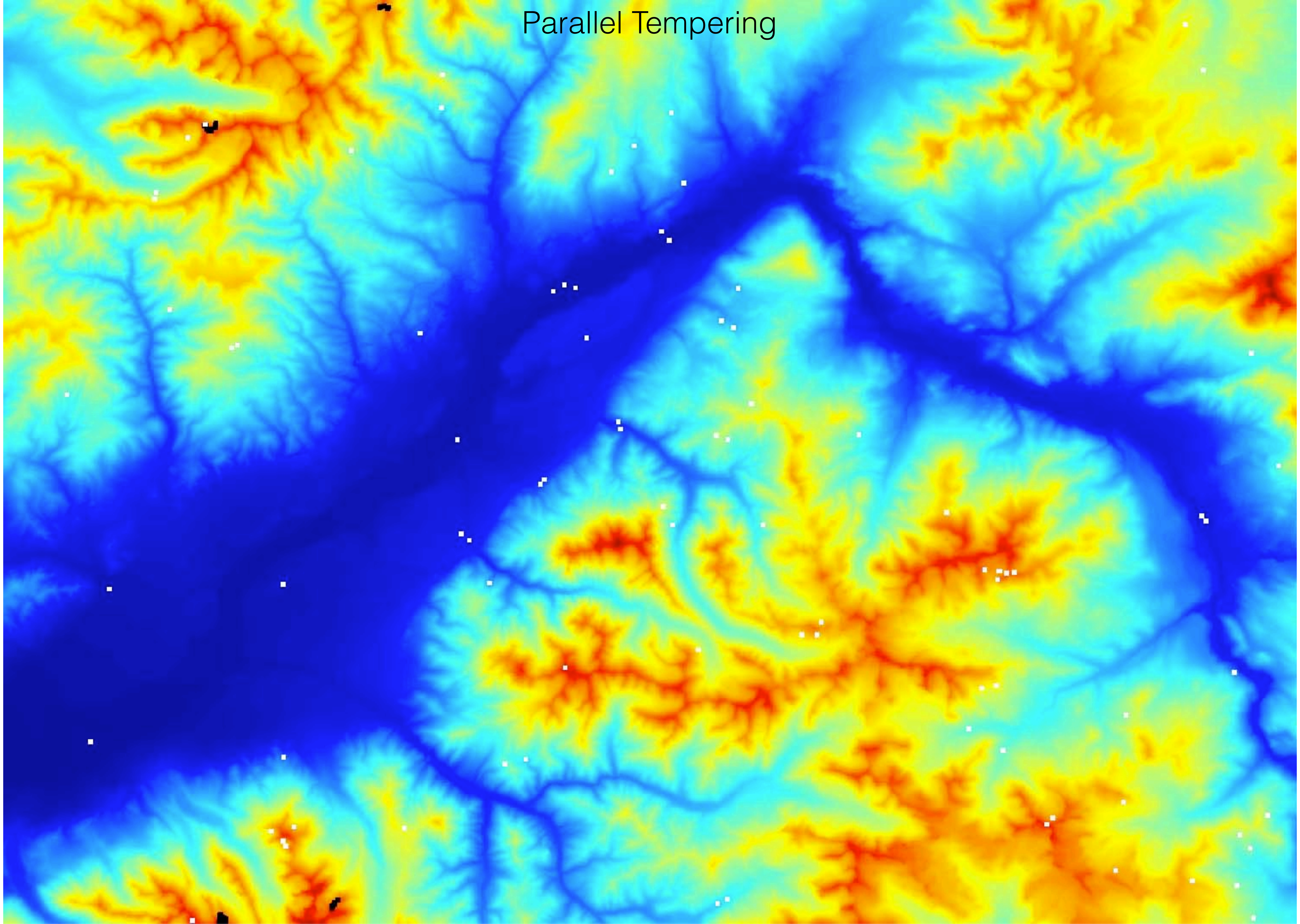
$$\log p(\mathbf{d}) = \int_0^1 \mathbb{E}[\log p(\mathbf{d}|\vec{\lambda})]_\beta \, d\beta$$

(Here $\beta = \frac{1}{T}$)

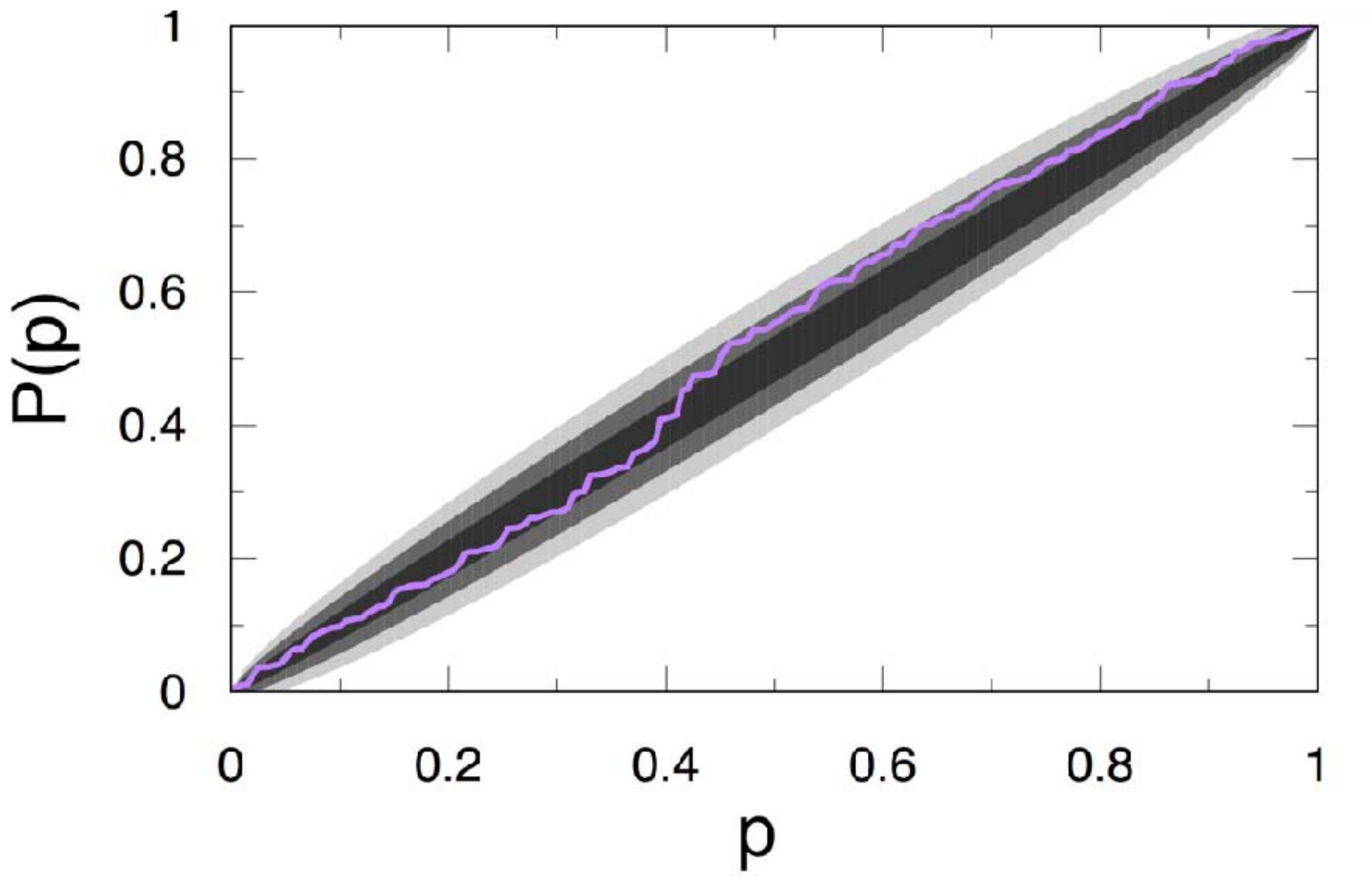
MCMC



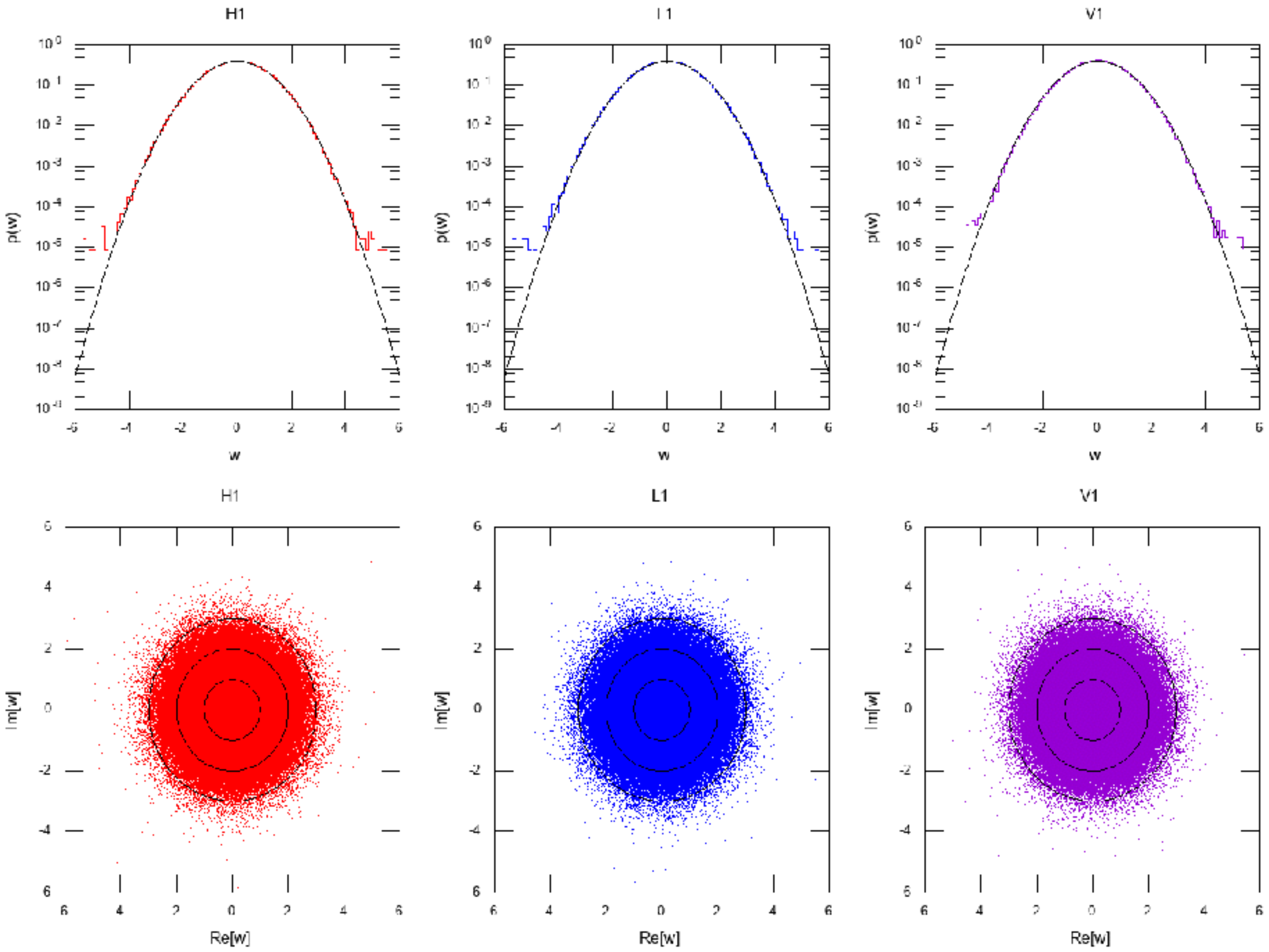
Parallel Tempering



Posterior Predictive Checks

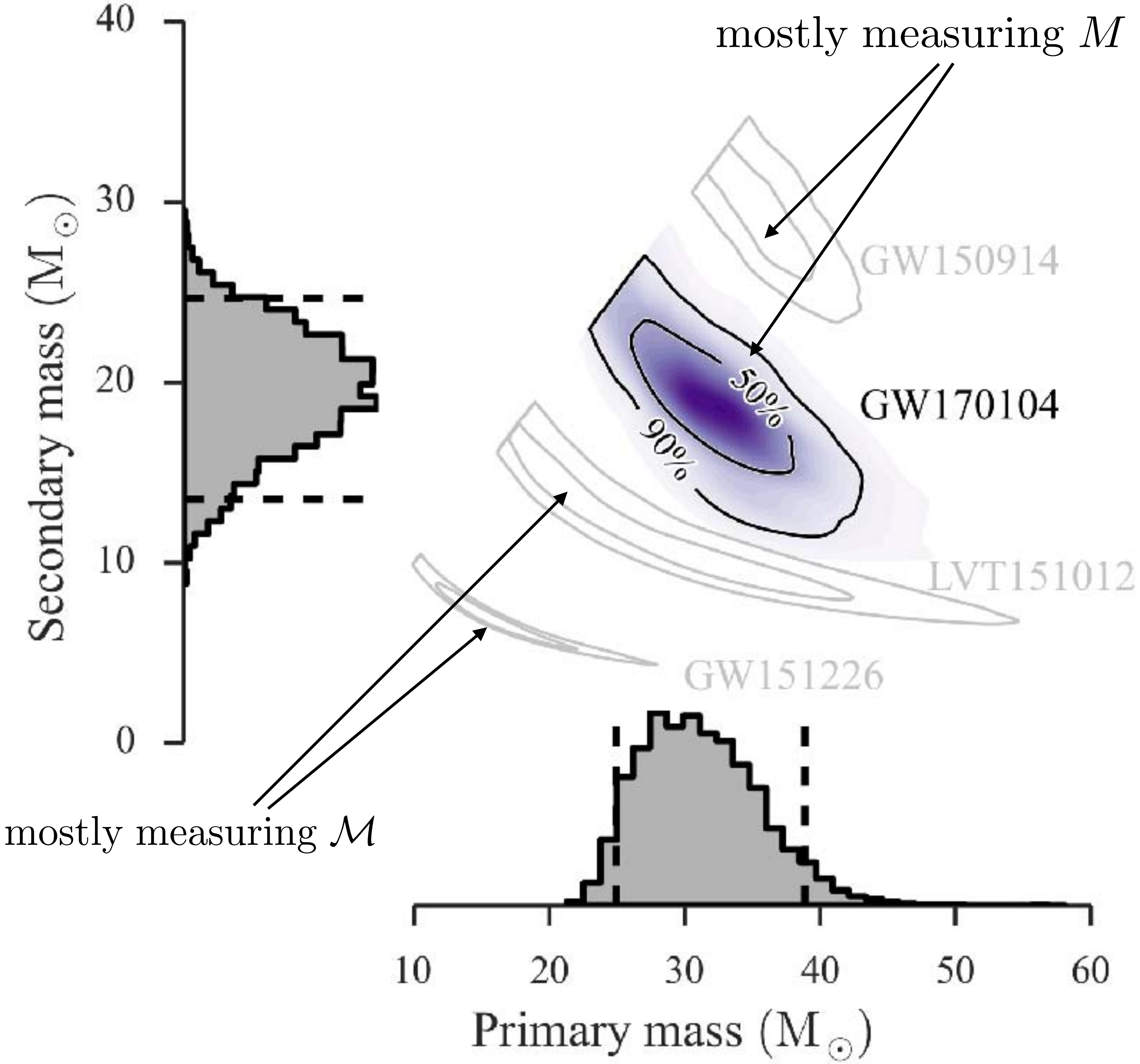


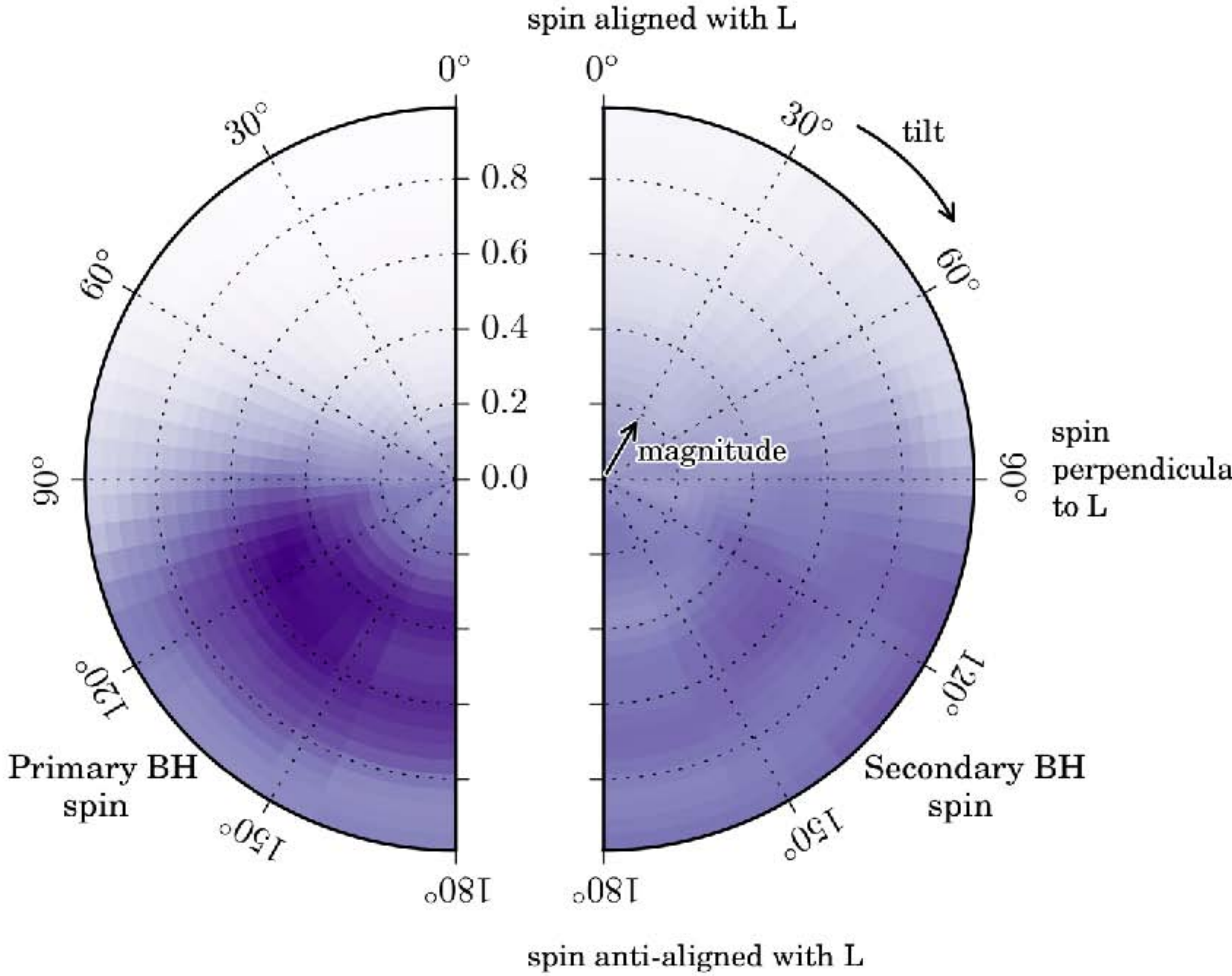
P-P plots, BW Sky localization



Gaussianity of BW residuals (noise model)

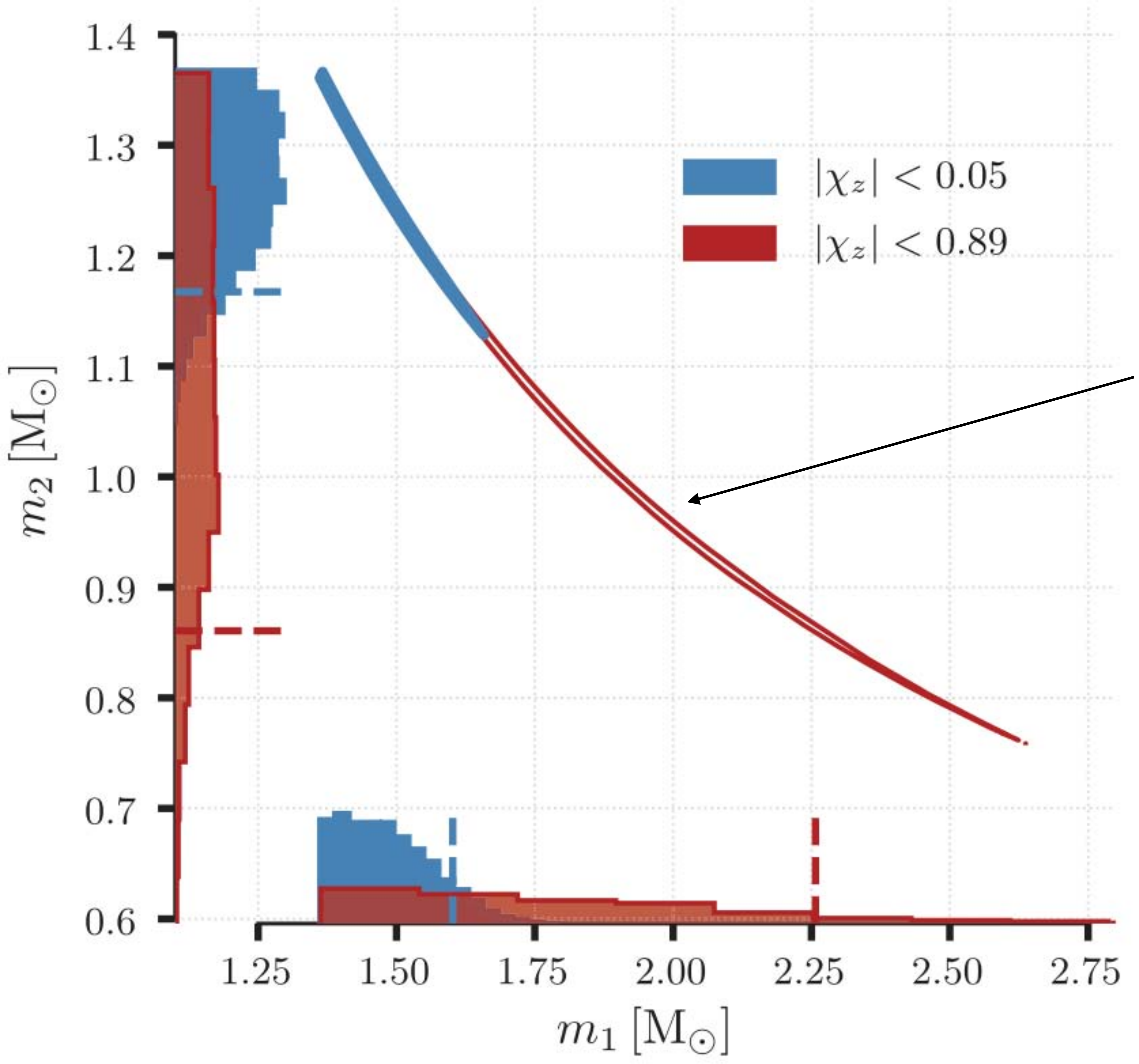
Parameter estimation: Examples from LIGO/Virgo





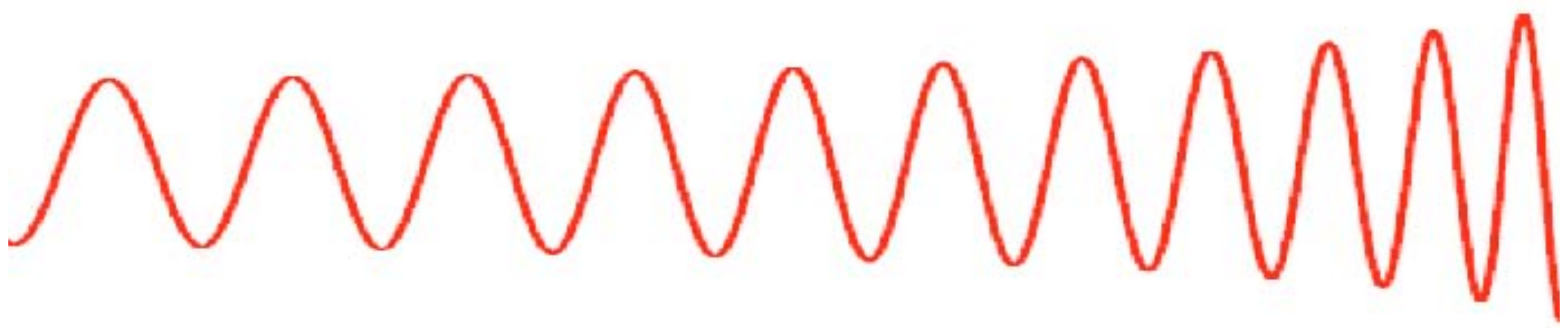
GW170104

BNS GW170817 - Parameter estimation



follows line of constant $\mathcal{M}$

Why measuring BH spin is hard I: Information from the Phase



$h(t)$

$$h(f) = \mathcal{A}_\ell(f) e^{i\Psi_\ell(f)}$$

Dominant Harmonic

$$A_2(f) = \frac{\mathcal{M}^2}{D_L\, u^{7/2}} \sum_{k=0} (\alpha_k(\vec{\lambda}) + \alpha_{lk}(\vec{\lambda}) \ln u) u^k$$

$$\Psi_2(f) = 2\pi f t_c - \Phi_c - \frac{\pi}{4} + \frac{3}{128\, u^5} \sum_{k=0} (\psi_k(\vec{\lambda}) + \psi_{lk}(\vec{\lambda}) \ln u) u^k$$

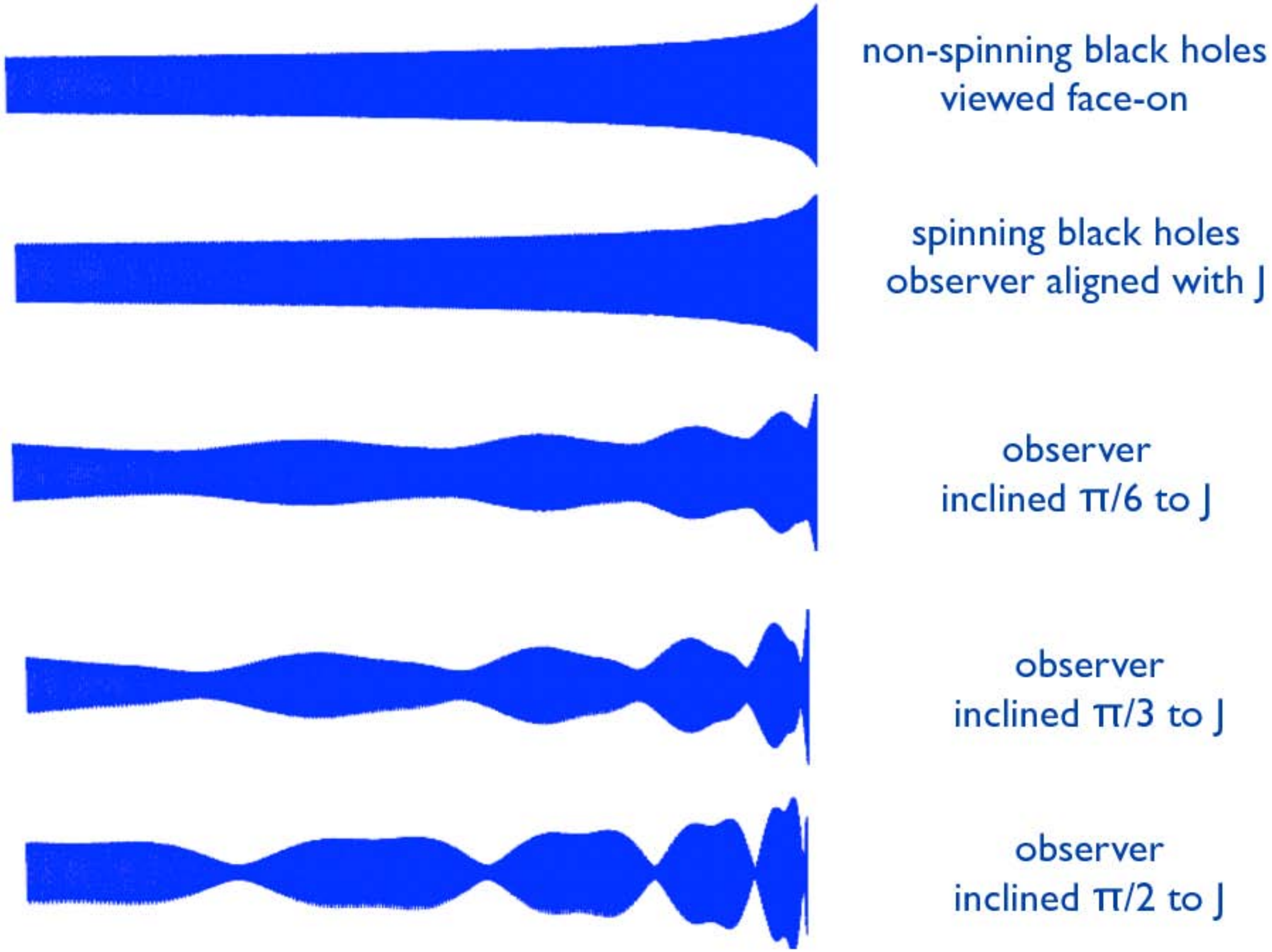
$$u = (\pi \mathcal{M} f)^{1/3} \sim v$$

Post-Newtonian Expansion

$$u = (\pi \mathcal{M} f)^{1/3} \sim v$$

0PN	$\frac{3}{128}\, u^{-5}$	Measure chirp mass
1PN	$\left(\frac{3715}{32256} + \eta \frac{55}{384}\right) \eta^{-2/5} u^{-3}$	Measure individual masses
1.5PN	$-\left(\frac{3\pi}{8} - \frac{1}{32}\left[113(1 \pm \sqrt{1-4\eta}) - 76\eta\right] \hat{L} \cdot \vec{\chi}_{1,2}\right) \eta^{-3/5}\, u^{-2}$	Measure spin combination
2PN	$\left(\frac{15293365}{21676032} + \frac{27145}{21504}\eta + \frac{3085}{3072}\eta^2 + \sigma(\hat{L} \cdot \vec{\chi}_{1,2},\, \vec{\chi_1} \cdot \vec{\chi_2},\, \chi_{1,2}^2)\right) \eta^{-4/5}\, u^{-1}$	Measure individual spins

Why measuring BH spin is hard II: Information from precision



Out-of-plane spin combination

$$\chi_{\rm eff} = m_1\,\chi_1\,\cos\theta_{LS_1} + m_2\,\chi_2\,\cos\theta_{LS_2}$$

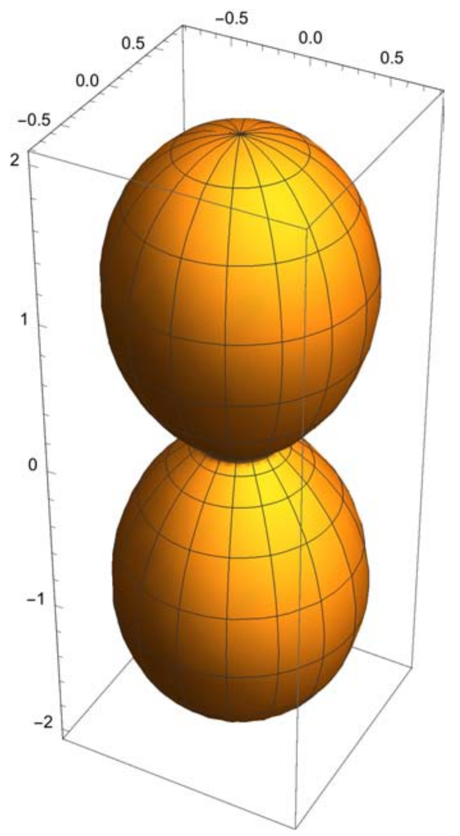
[component (anti)aligned with angular momentum]

In-plane spin combination

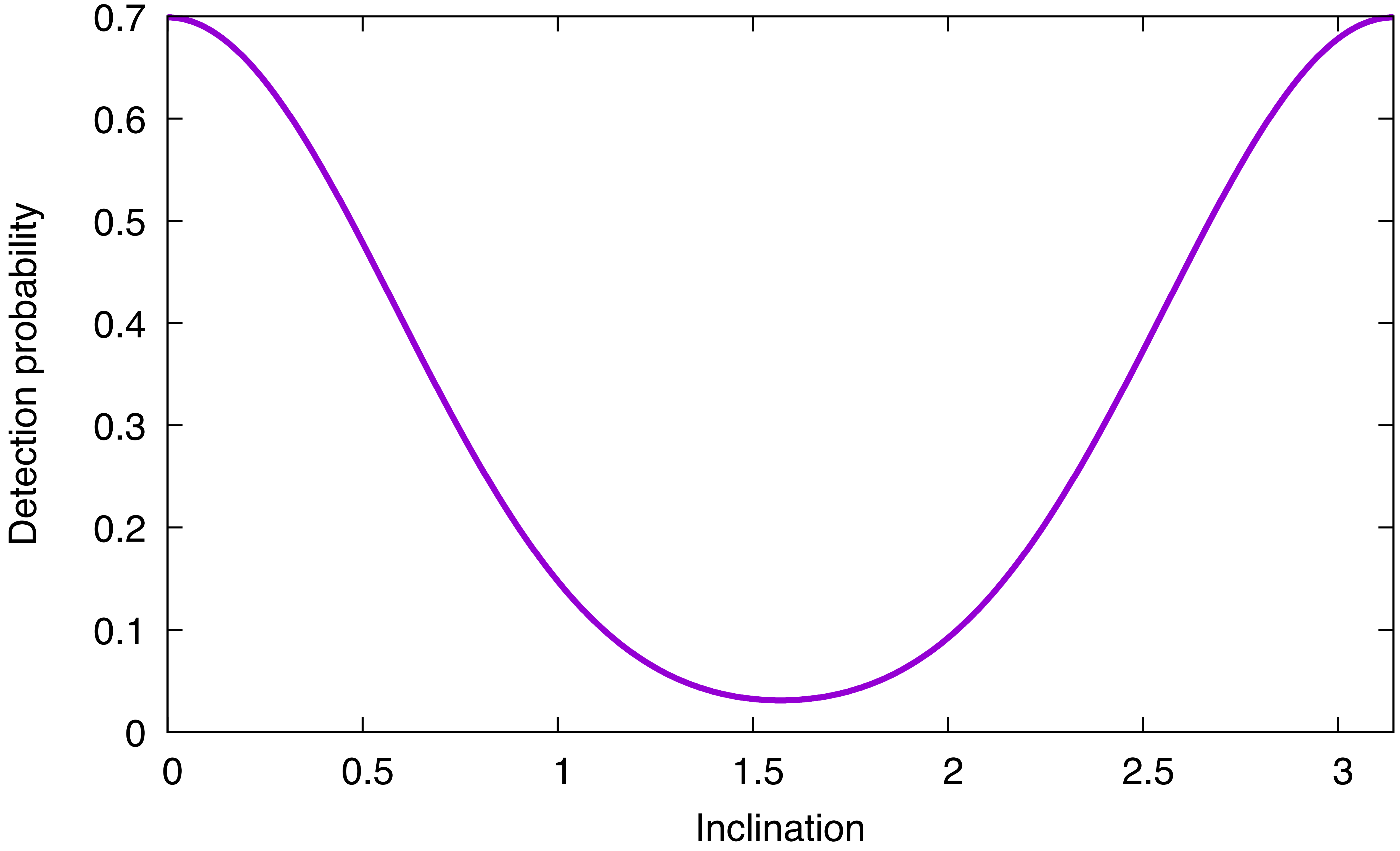
$$\chi_{\rm p} = \frac{1}{2}\left(\chi_{2\perp} + \alpha\,\chi_{1\perp} + |\chi_{2\perp} - \alpha\,\chi_{1\perp}|\right)$$

$$\alpha = \left(\frac{m_1}{m_2}\right)\frac{(4M - m_2)}{(4M - m_1)}$$

Selection Effects



Face-on Edge-on Face-off

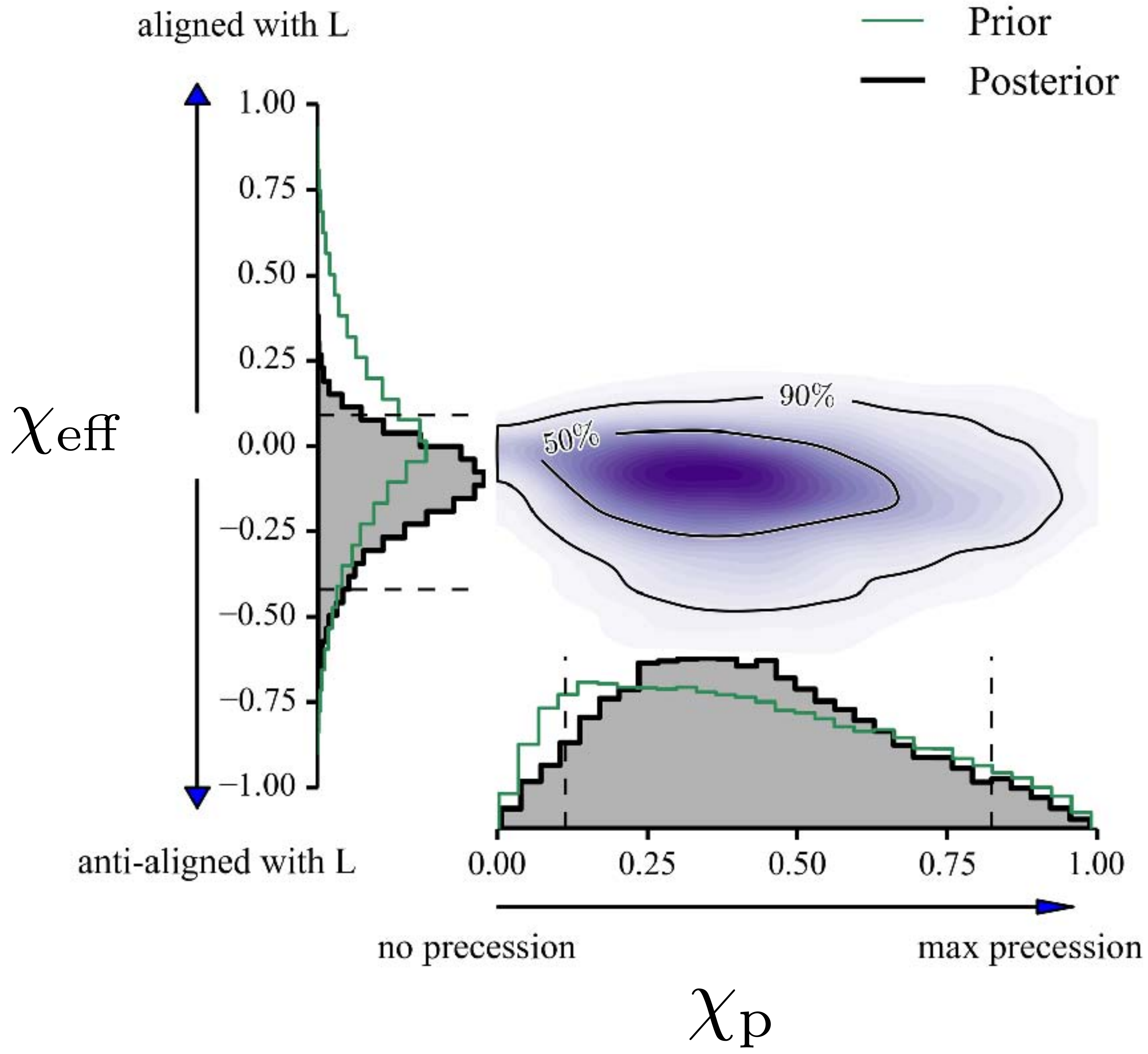


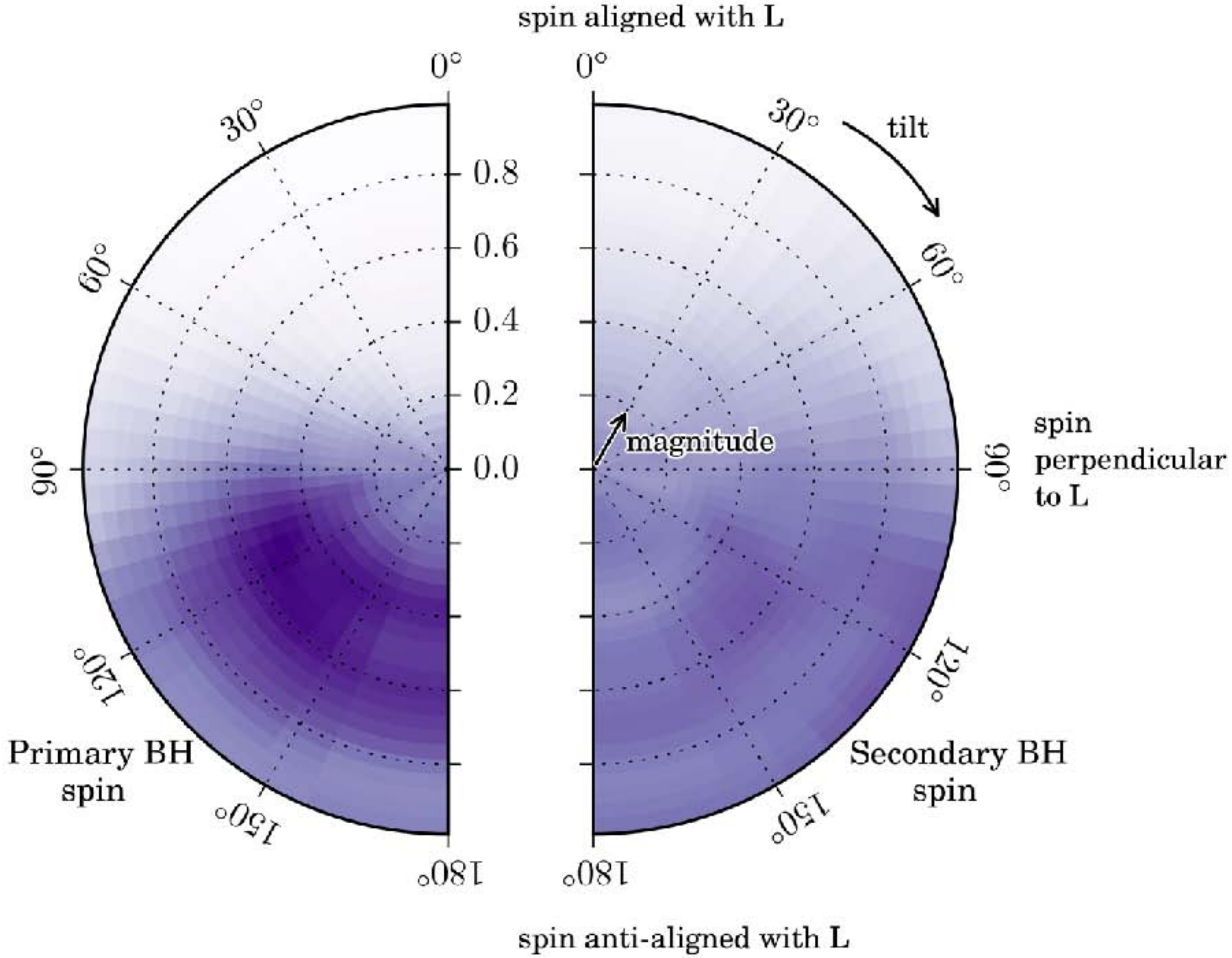
We are more likely to detect
Face-on/off systems than
Edge-on systems
Harder to measure precession

We are also more likely to
detect more massive systems
70 Mpc for 1.4+1.4 Msun
300 Mpc for 10+10 Msun
700 Mpc for 30+30 Msun

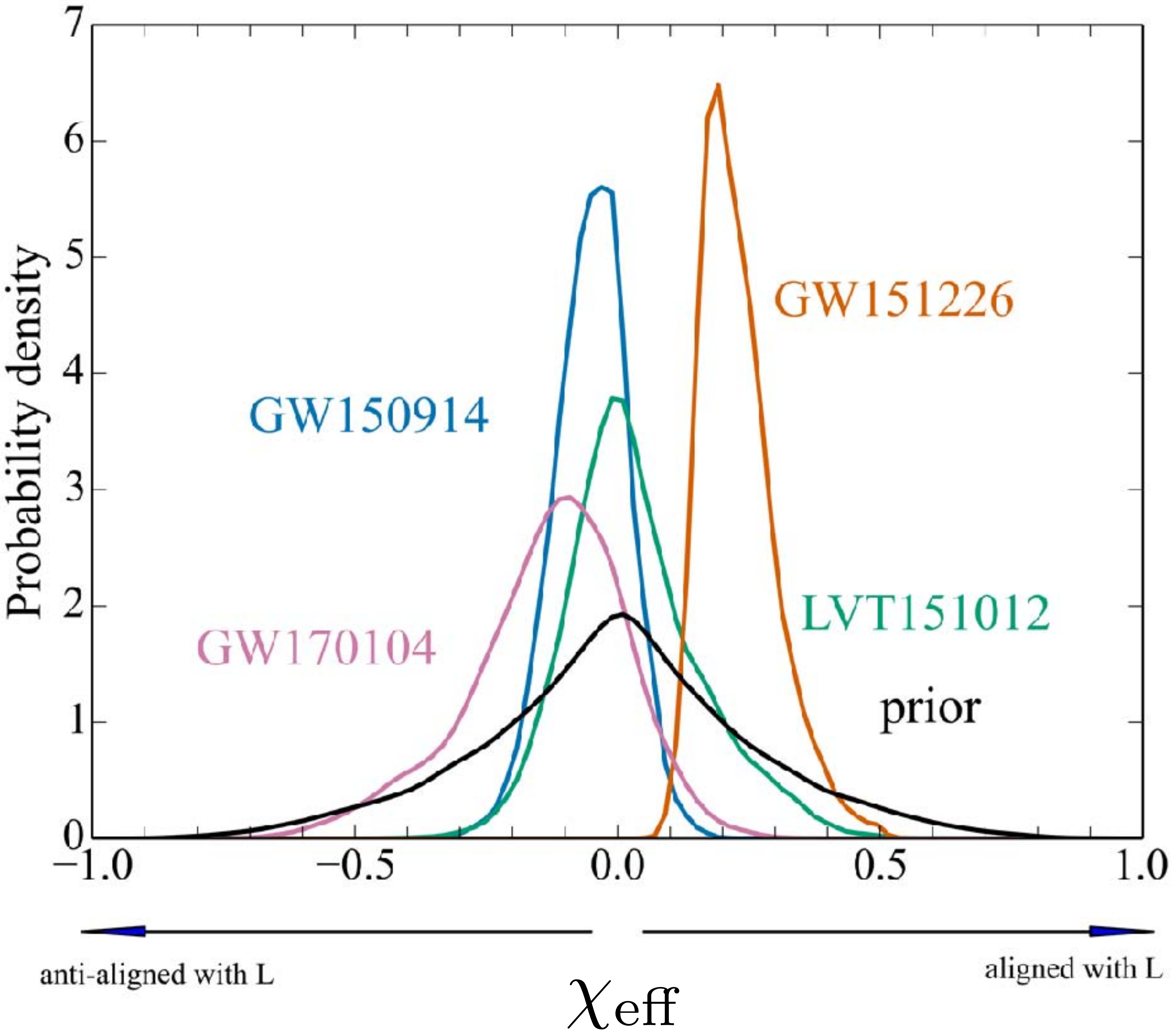
More massive, less cycles in-band,
Harder to measure precession

Spin posteriors for GW170104





Evidence for spin precession remains elusive



All four events consistent with low spin, no precession

$|\chi_{\rm eff}| < 0.35$ 90% confidence, all 4

$\chi_p \sim 0$