Gravitational wave data analysis

Neil J. Cornish

extreme Gravity Institute

MONTANA STATE UNIVERSITY

Resources

Papers/Reviews

LIGO/Virgo, “A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals”, CQG **37**, 055002 (2020)

Cornish “Black Hole Merging and Gravitational Waves”, Black Hole Formation and Growth, Saas-Fee Advanced Course 48, (2019)
https://www.dropbox.com/s/l8nusg5fd5x3ak1/2019_Book_BlackHoleFormationAndGrowth.pdf?dl=0

Romano & Cornish “Detection methods for stochastic gravitational-wave backgrounds: a unified treatment” Living Rev.Rel. **20, 1** (2017)

Books

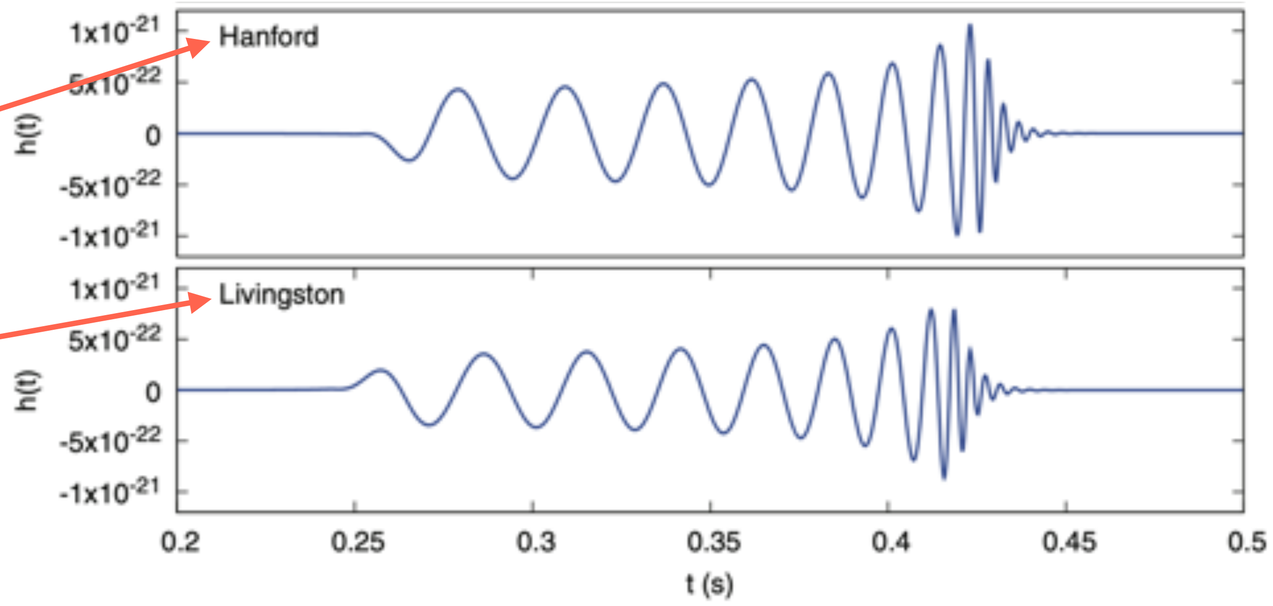
Maggiore, “Gravitational Waves: Volume 1: Theory and Experiments”

Creighton & Anderson “Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis”

Outline of lectures

- Detector Response Functions
- Source localization
- Data Analysis 101 - the Likelihood function
- Statistical Framework, Bayesian and Frequentist
- Searching for signals
- Bayesian Inference, parameter estimation
- Transdimensional inference
- Noise modeling
- Recent developments

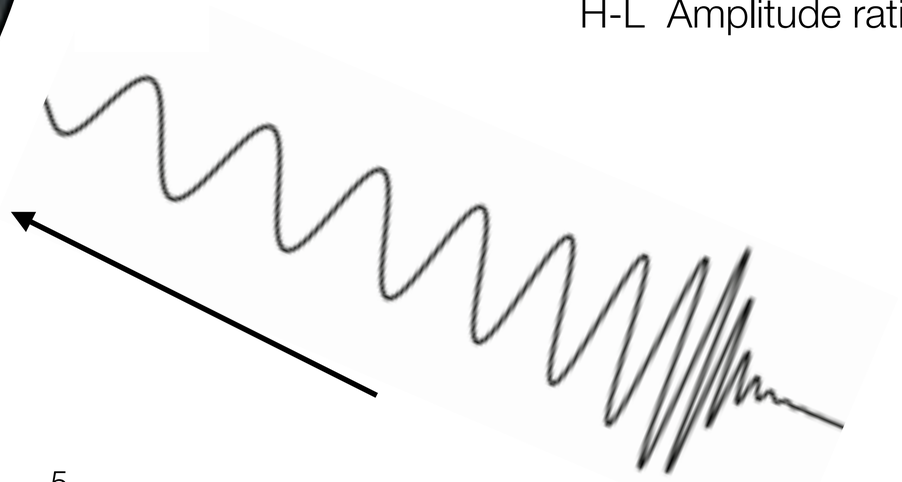

Scale of Effect Vastly Exaggerated



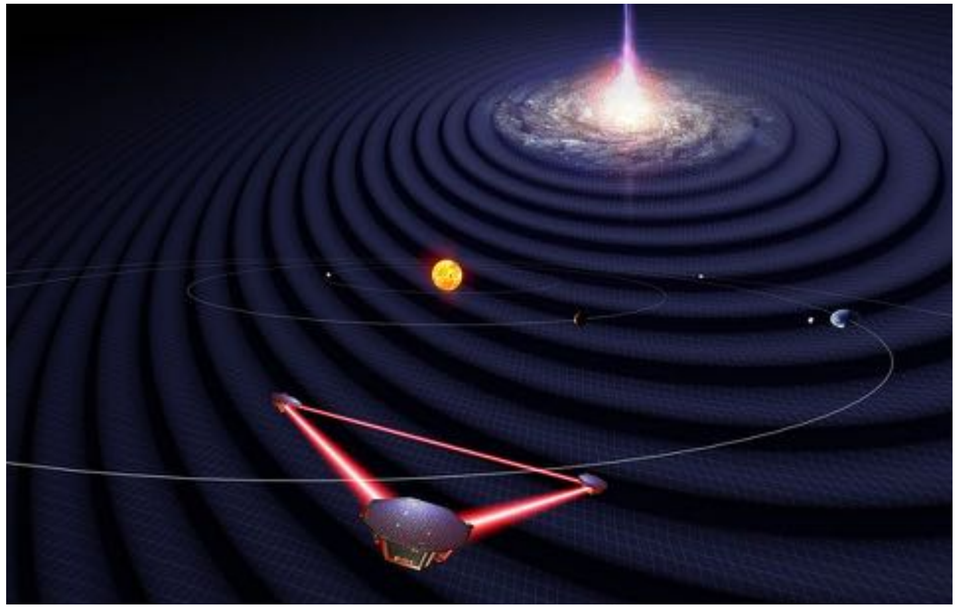
H-L Time delay 7 ms

H-L Phase Shift 2.9 radians

H-L Amplitude ratio 1.24

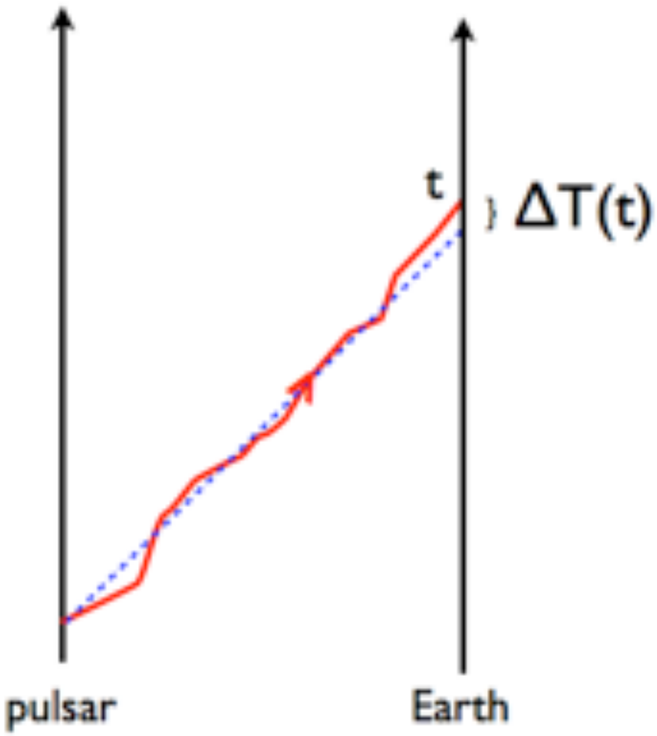


Gravitational Wave Detectors - Time of Flight



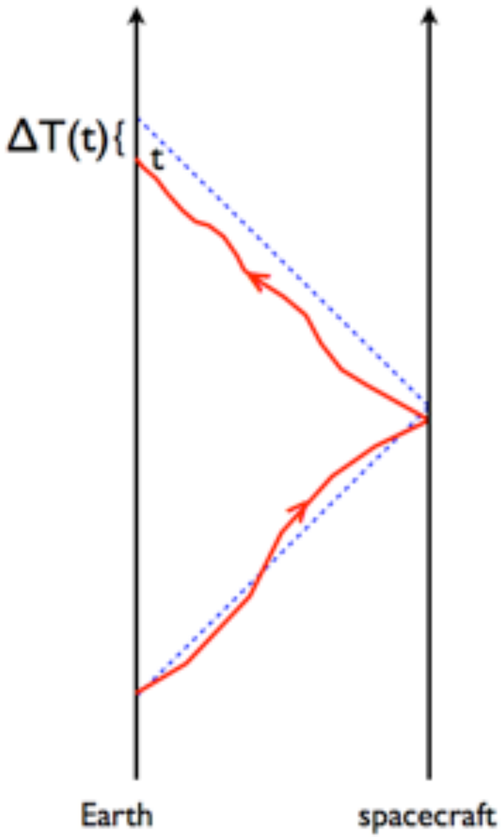
Detector Response to GWs

$$\Delta T(t)$$



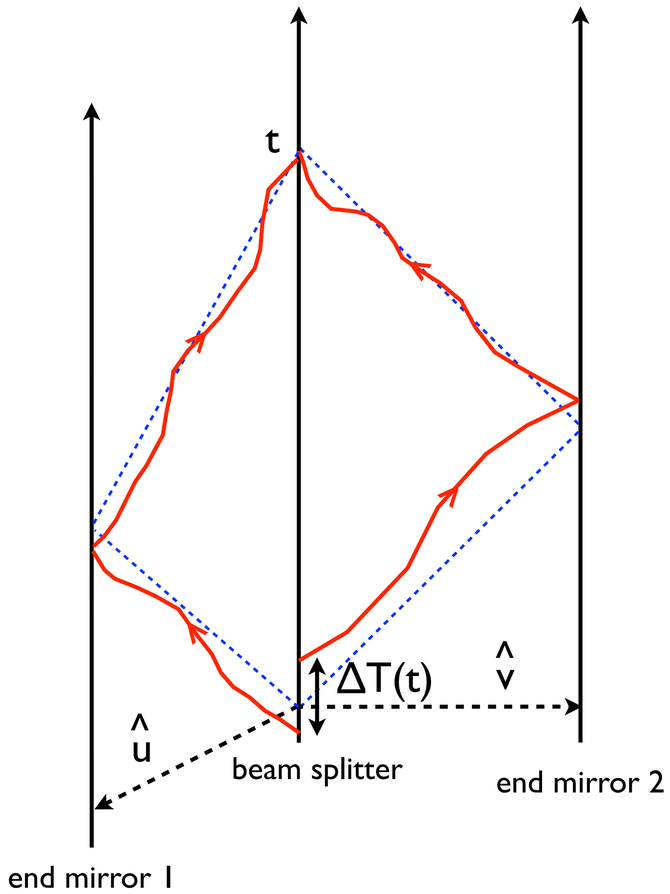
Pulsar Timing

$$\frac{\Delta\nu(t)}{\nu_0} = \frac{d\Delta T(t)}{dt}$$



Spacecraft tracking

$$\Phi(t) = 2\pi\nu_0\Delta T(t)$$



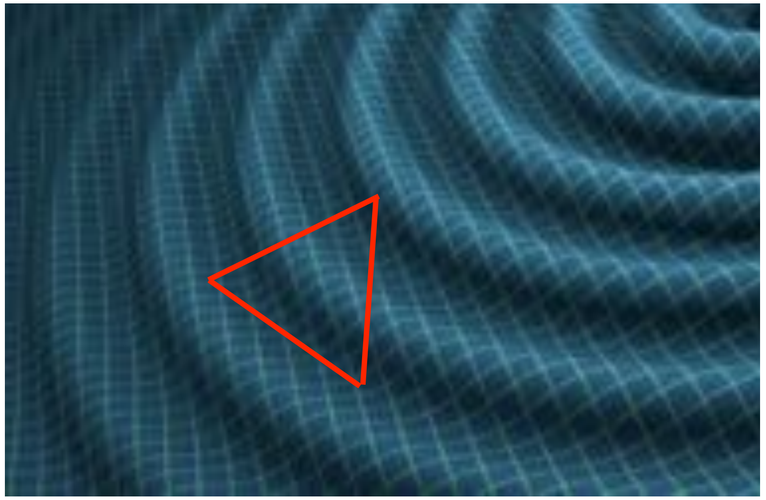
Laser Interferometers

The Long and the Short of it

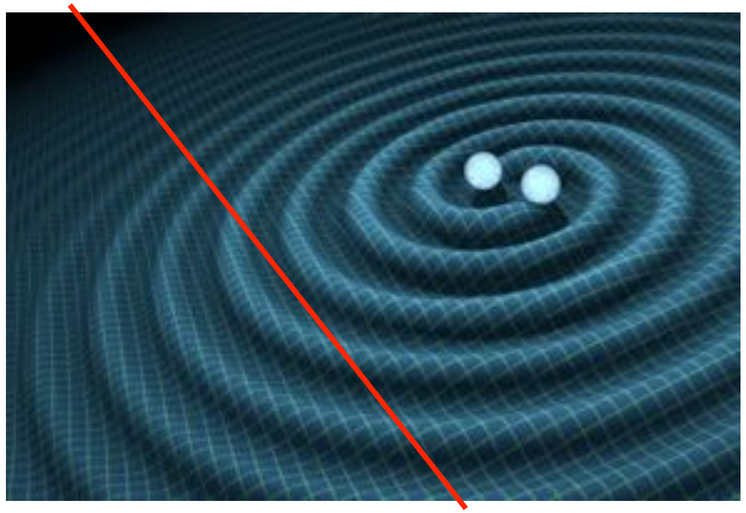
Beam detector	L (km)	f_* (Hz)	f (Hz)	f/f_*	Relation
Ground-based interferometer	~ 1	$\sim 10^5$	10 to 10^4	10^{-4} to 10^{-1}	$f \ll f_*$
Space-based interferometer	$\sim 10^6$	$\sim 10^{-1}$	10^{-4} to 10^{-1}	10^{-3} to 1	$f \lesssim f_*$
Spacecraft Doppler tracking	$\sim 10^9$	$\sim 10^{-4}$	10^{-6} to 10^{-3}	10^{-2} to 10	$f \sim f_*$
Pulsar timing	$\sim 10^{17}$	$\sim 10^{-12}$	10^{-9} to 10^{-7}	10^3 to 10^5	$f \gg f_*$

$$f_* = \frac{c}{L}$$

LIGO



LISA



PTA

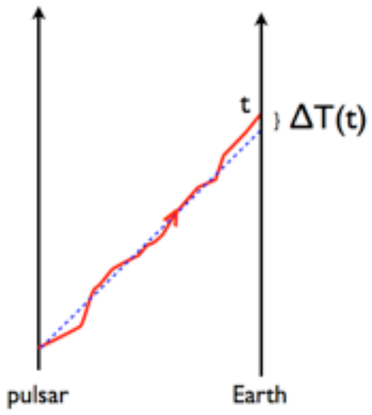
Time of flight computed in TT gauge

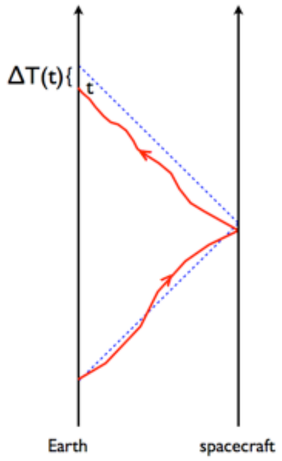
$$ds^2 = -dt^2 + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx + dz^2$$

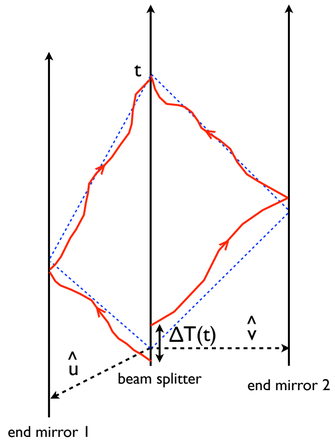
$$= -dvdu + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$

where $u = t - z, \quad v = t + z$

All time-of-flight detectors require us to compute the time it takes a photon to travel from one event to another in the spacetime perturbed by a GW. Some require multiple trips







Time of flight computed in TT gauge

$$ds^2 = -dudv \;\; + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$

Have to solve for null geodesics in this metric. We could integrate the geodesic equations, but the spacetime has lots of symmetry, and hence conserved quantities. No integration needed!

Killing vectors $\vec{\partial}_x,\ \vec{\partial}_y,\ \vec{\partial}_v$ Photon worldline $x^\alpha(\lambda)$ Photon 4-velocity $S^\alpha = \dfrac{dx^\alpha}{d\lambda}$

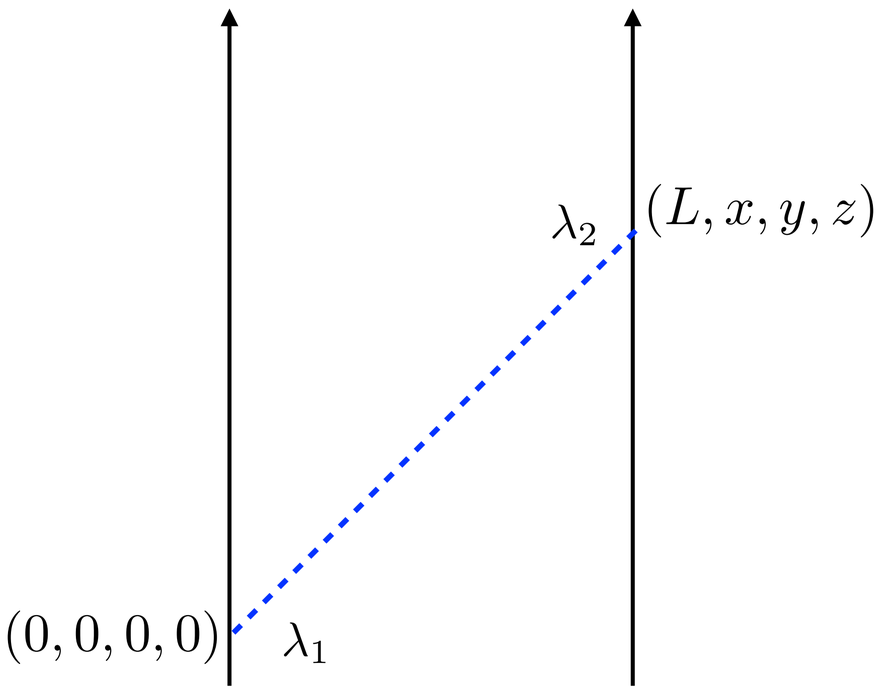
Killing vectors yield three constants of motion $S_x,\ S_y,\ S_v$ and the condition $S_\alpha S^\alpha = 0$

[Derivation follows N. J. Cornish, Phys.Rev.D80:087101,2009, arXiv:0910.4372]

[Based on F. B. Estabrook and H. D. Wahlquist, Gen. Rel. Grav. 6, 439 (1975)]

Time of flight computed in TT gauge

$$ds^2 = -dudv \;\; + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$



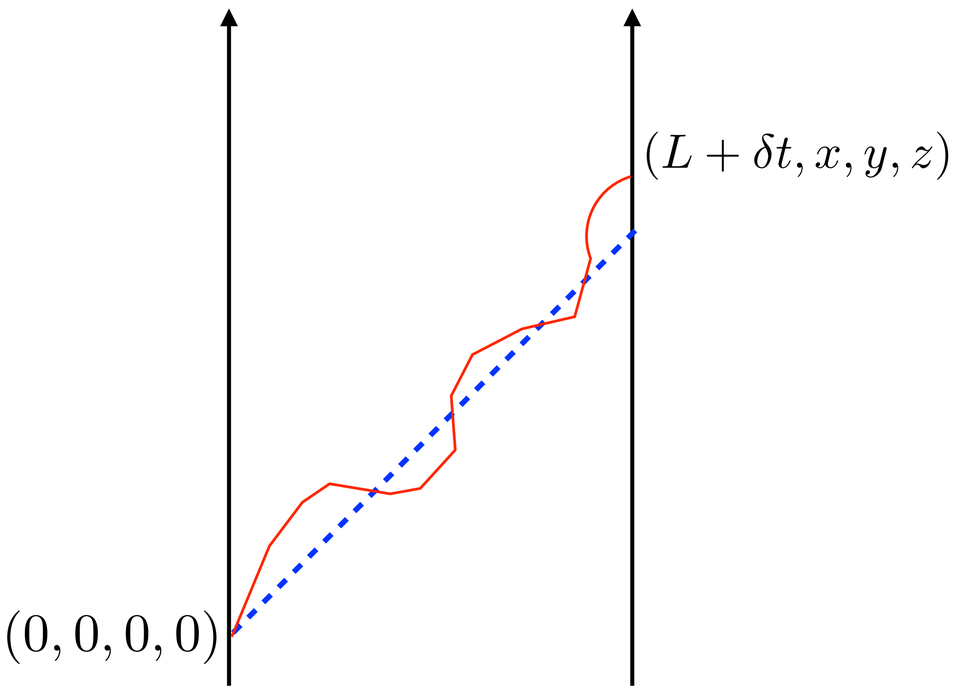
Path from $(0, 0, 0, 0)$ to (L, x, y, z) in unperturbed spacetime has

$$S_x = \frac{x}{\Delta\lambda}, \quad S_y = \frac{y}{\Delta\lambda}, \quad S_v = \frac{z - L}{2\Delta\lambda}$$

$$S_\alpha S^\alpha = 0 \qquad \Rightarrow \qquad t = L = \sqrt{x^2 + y^2 + z^2}$$

Time of flight computed in TT gauge

$$ds^2 = -dudv^2 + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$



When GWs are present the trajectory is perturbed

$$x^\mu(\lambda) = x_0^\mu(\lambda) + \delta x^\mu(\lambda)$$

Spatial endpoints fixed TT gauge: $\delta x^i(0) = \delta x^i(\Delta\lambda) = 0$

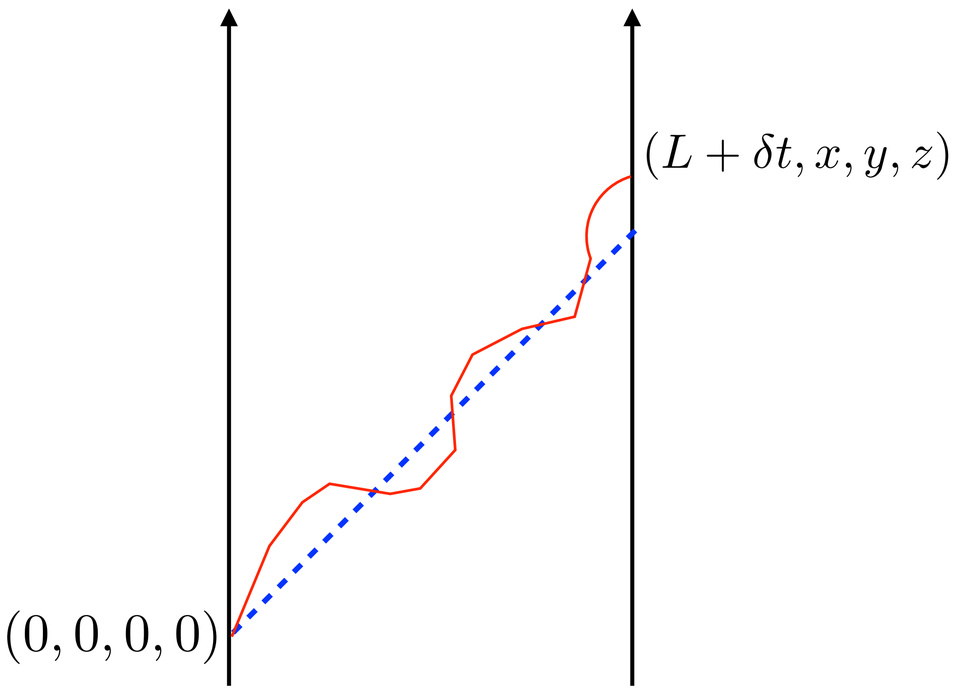
Conserved quantities are:

$$S_x = (1 + h_+)u^x + h_\times u^y$$
$$S_y = (1 - h_+)u^y + h_\times u^x$$
$$S_v = -\frac{1}{2}u^u$$
$$0 = \alpha_x u^x + \alpha_y u^y + 2\alpha_v u^v$$

Expand and solve for perturbations: $S_\mu \to S_\mu^0 + \delta S_\mu$

Time of flight computed in TT gauge

$$ds^2 = -dudv^2 + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$



For example: $S_x = (1 + h_+)u^x + h_\times u^y$

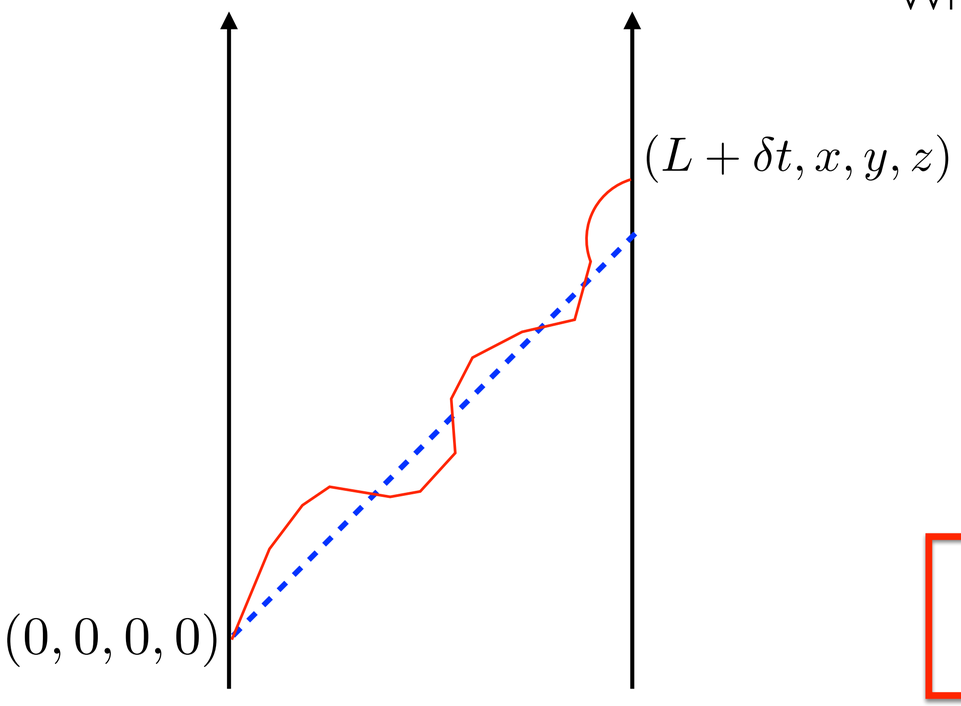
Becomes $\delta S_x = \dfrac{d\delta x}{d\lambda} + h_+ u_0^x + h_\times u_0^y$

$$\Rightarrow \quad \delta S_x \Delta\lambda = \delta x(\Delta\lambda) - \delta x(0) + u_0^x \int h_+ \, d\lambda + u_0^y \int h_\times \, d\lambda$$

$$\Rightarrow \quad \delta S_x = \frac{x}{u^u \Delta\lambda^2} \int h_+ \, du + \frac{y}{u^u \Delta\lambda^2} \int h_\times \, du$$

Time of flight computed in TT gauge

$$ds^2 = -dudv^2 + dx^2(1 + h_+(u)) + dy^2(1 - h_+(u)) + 2h_\times(u)dydx$$



When GWs are present we have to change our aim:

$$\delta S_x = \frac{xH_{xx} + yH_{xy}}{\Delta\lambda(L - z)}$$

$$\delta S_y = \frac{yH_{yy} + xH_{xy}}{\Delta\lambda(L - z)}$$

$$\delta S_v = -\frac{\delta t}{2\,\Delta\lambda}$$

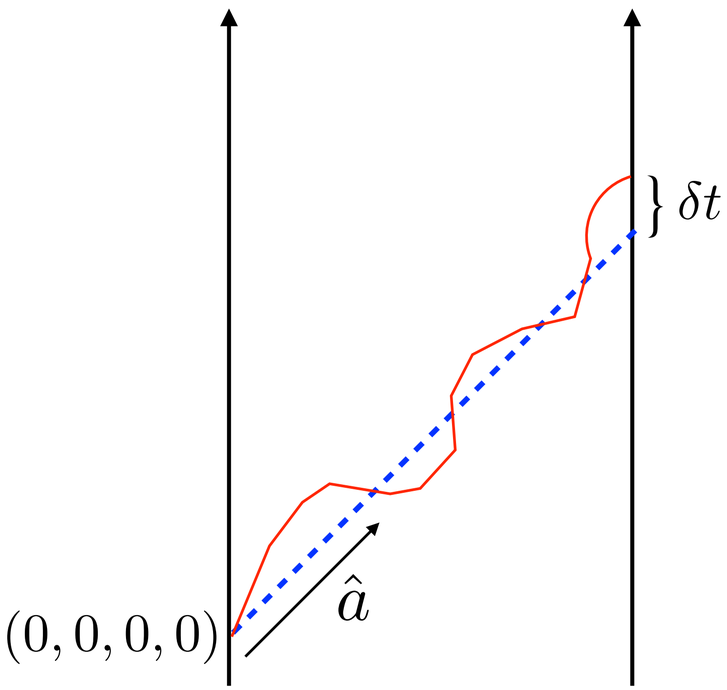
$$H_{ij} = \int_0^{L-z} h_{ij}(u)du$$

$$\boxed{\delta t = \frac{1}{2L(L - z)}(x^2 H_{xx} + y^2 H_{yy} + 2xyH_{xy})}$$

$$(h_{xx} = -h_{yy} = h_+, \quad h_{xy} = h_\times)$$

Time of flight computed in TT gauge

$$\delta t = \frac{1}{2L(L-z)}(x^2 H_{xx} + y^2 H_{yy} + 2xyH_{xy})$$



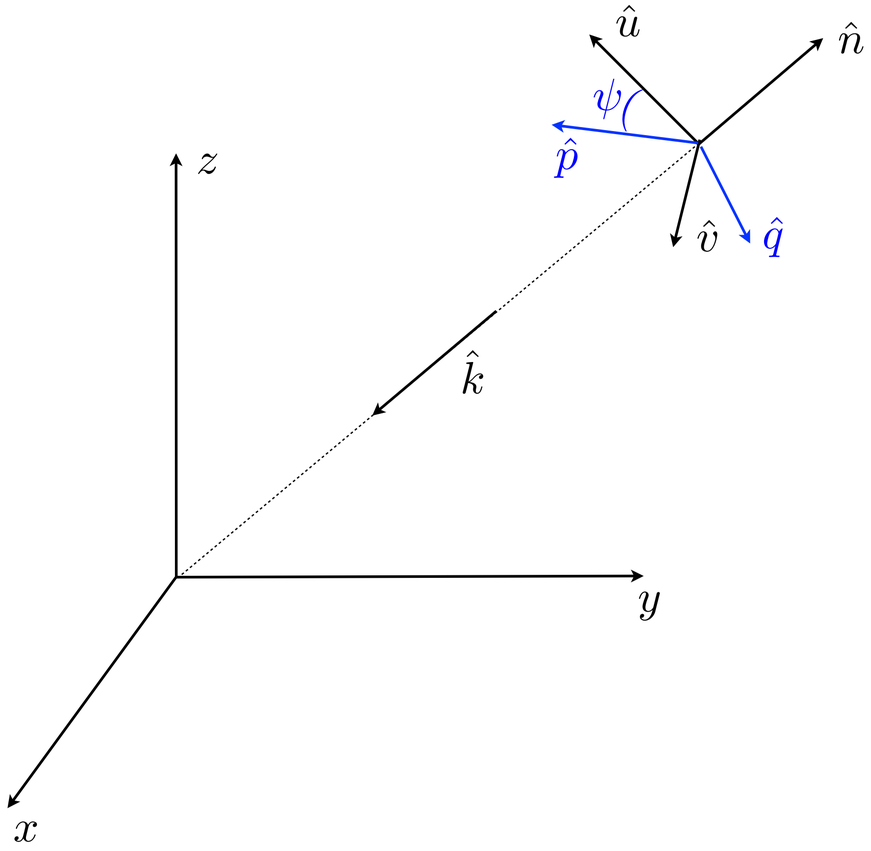
Coordinate independent version:

$$\Delta\tau_{12} = \frac{(\hat{a}\otimes\hat{a}) : \mathbf{H}[u_1, u_2]}{2(1-\hat{k}\cdot\hat{a})} \qquad (u = k_\alpha x^\alpha)$$

Here $\hat{a}$ is a unit vector along the detector arm and $\hat{k}$ is the GW propagation direction

$$\mathbf{H}[u_1, u_2] = \int_{u_1}^{u_2} \mathbf{h}(u)\, du \qquad \mathbf{h} = h_+(u)\,\epsilon^+ + h_\times(u)\,\epsilon^\times$$

General coordinate system



$$\hat{n} = \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z}$$
$$\hat{u} = \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z}$$
$$\hat{v} = \sin\phi\,\hat{x} - \cos\phi\,\hat{y}$$

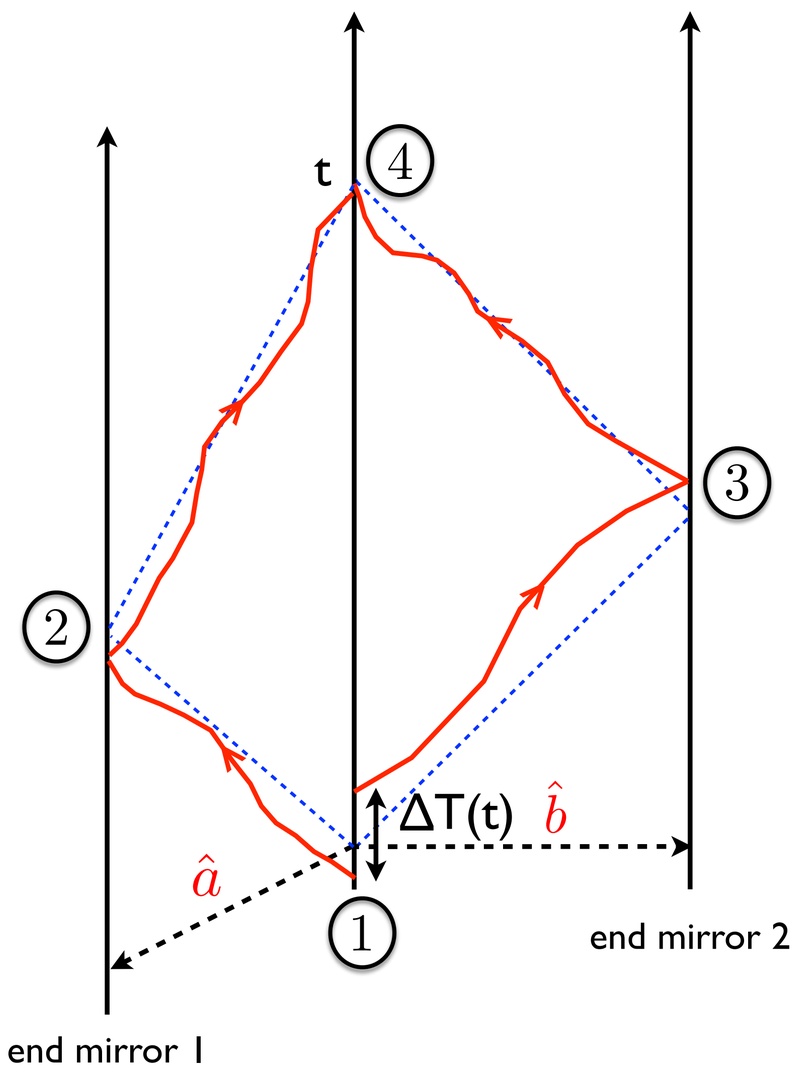
$$\mathbf{e}^{+} = \hat{u}\otimes\hat{u} - \hat{v}\otimes\hat{v}$$
$$\mathbf{e}^{\times} = \hat{u}\otimes\hat{v} + \hat{v}\otimes\hat{u}$$

$$\mathbf{h} = h_{+}\boldsymbol{\epsilon}^{+} + h_{\times}\boldsymbol{\epsilon}^{\times}$$

$$\begin{aligned}\boldsymbol{\epsilon}^{+} &= \hat{p}\otimes\hat{p} - \hat{q}\otimes\hat{q} \\ &= \cos 2\psi\,\mathbf{e}^{+} - \sin 2\psi\,\mathbf{e}^{\times}\end{aligned}$$

$$\begin{aligned}\boldsymbol{\epsilon}^{\times} &= \hat{p}\otimes\hat{q} + \hat{q}\otimes\hat{p} \\ &= \sin 2\psi\,\mathbf{e}^{+} + \cos 2\psi\,\mathbf{e}^{\times}\end{aligned}$$

Example: Laser interferometer in the long wavelength limit



$$\Delta T(t) = \Delta\tau_{12} + \Delta\tau_{24} - \Delta\tau_{13} - \Delta\tau_{34}$$

$$h(t) \equiv \frac{\Delta T(t)}{2L} \approx \underbrace{\frac{1}{2}\left[\hat{a}\otimes\hat{a} - \hat{b}\otimes\hat{b}\right]}_{\text{Detector tensor}} : \mathbf{h}(t)$$

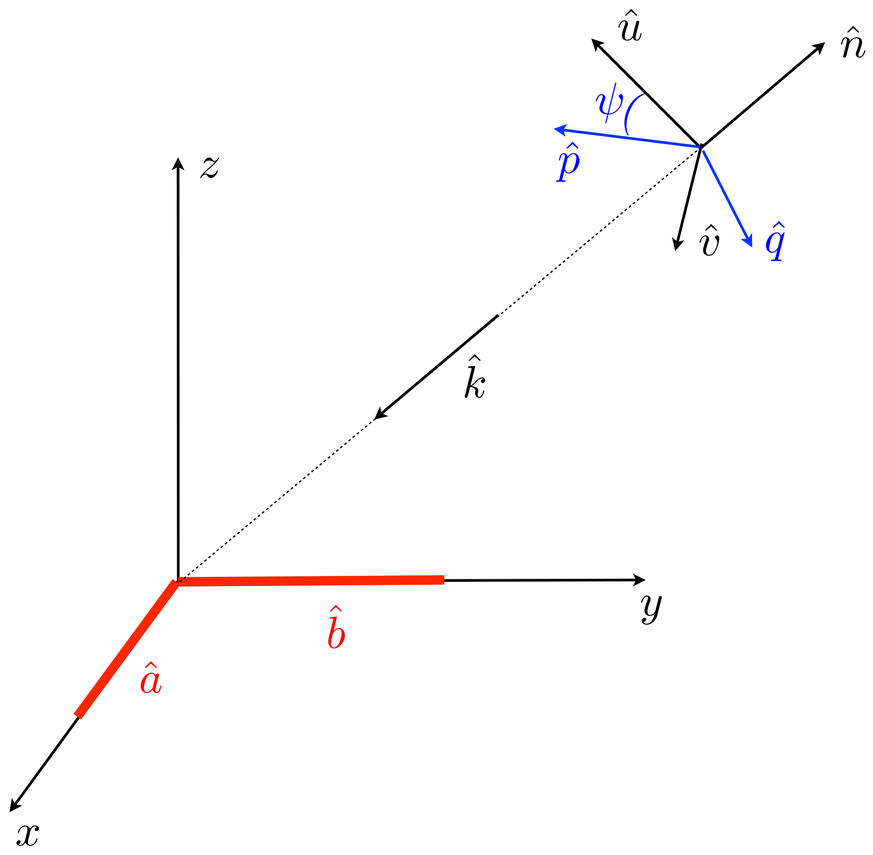
$$\mathbf{h}(t) = h_+(t)\boldsymbol{\epsilon}^+ + h_\times(t)\boldsymbol{\epsilon}^\times$$

Polarization tensors

Antenna Pattern Functions

$$\hat{n} = \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z}$$
$$\hat{u} = \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z}$$
$$\hat{v} = \sin\phi\,\hat{x} - \cos\phi\,\hat{y}$$

$$\mathbf{e}^{+} = \hat{u}\otimes\hat{u} - \hat{v}\otimes\hat{v}$$
$$\mathbf{e}^{\times} = \hat{u}\otimes\hat{v} + \hat{v}\otimes\hat{u}$$



$$(\hat{a}\otimes\hat{a}) : \mathbf{e}^{+} = \cos^2\theta\cos^2\phi - \sin^2\phi$$

$$(\hat{a}\otimes\hat{a}) : \mathbf{e}^{\times} = \cos\theta\sin 2\phi$$

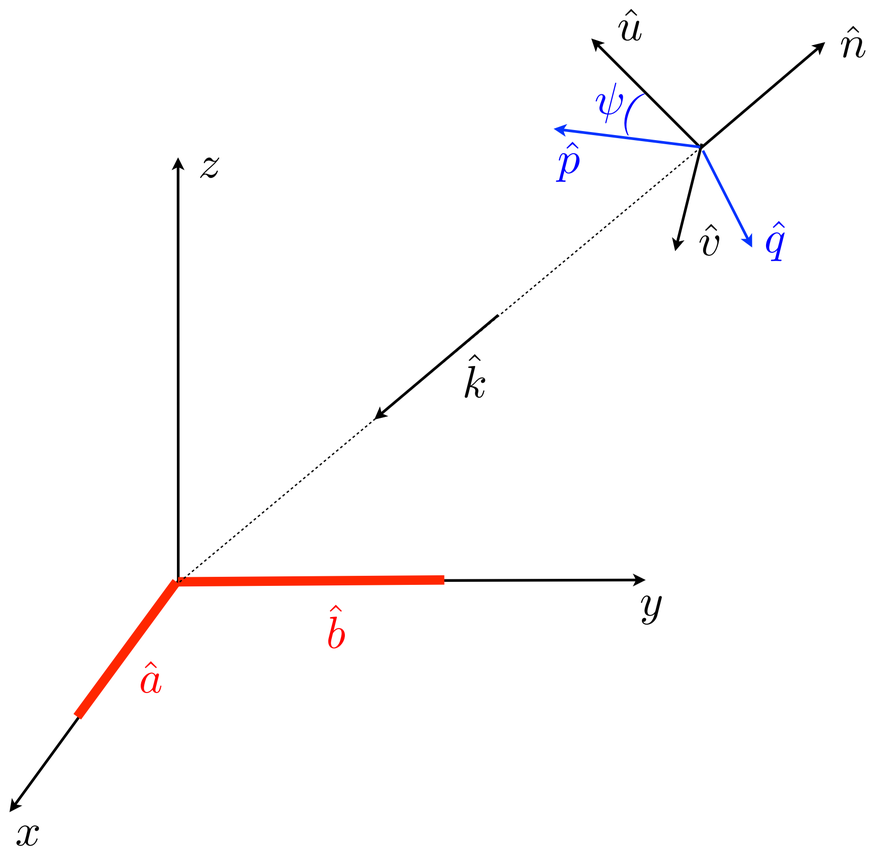
$$(\hat{b}\otimes\hat{b}) : \mathbf{e}^{+} = \cos^2\theta\sin^2\phi - \cos^2\phi$$

$$(\hat{b}\otimes\hat{b}) : \mathbf{e}^{\times} = -\cos\theta\sin 2\phi$$

Antenna Pattern Functions

$$\hat{n} = \sin\theta\cos\phi\,\hat{x} + \sin\theta\sin\phi\,\hat{y} + \cos\theta\,\hat{z}$$
$$\hat{u} = \cos\theta\cos\phi\,\hat{x} + \cos\theta\sin\phi\,\hat{y} - \sin\theta\,\hat{z}$$
$$\hat{v} = \sin\phi\,\hat{x} - \cos\phi\,\hat{y}$$

$$\mathbf{e}^{+} = \hat{u}\otimes\hat{u} - \hat{v}\otimes\hat{v}$$
$$\mathbf{e}^{\times} = \hat{u}\otimes\hat{v} + \hat{v}\otimes\hat{u}$$



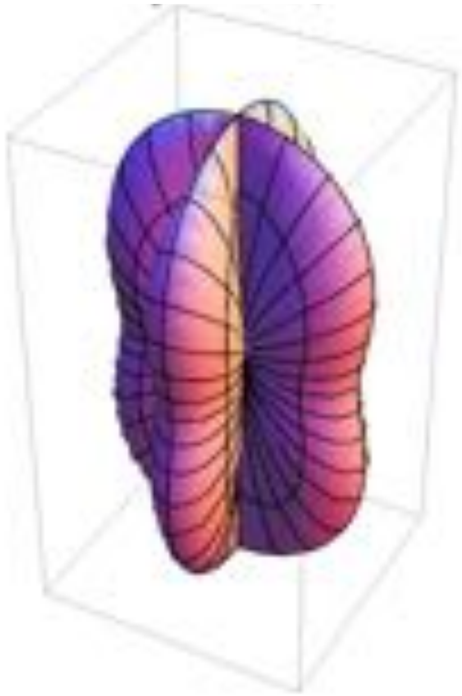
$$h = F^{+}h_{+} + F^{\times}h_{\times}$$

$$\begin{aligned} F^{+} &= \frac{1}{2}(\hat{a}\otimes\hat{a} - \hat{b}\otimes\hat{b}) : \epsilon^{+} \\ &= \frac{1}{2}(1+\cos^2\theta)\cos(2\phi)\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi \end{aligned}$$

$$\begin{aligned} F^{\times} &= \frac{1}{2}(\hat{a}\otimes\hat{a} - \hat{b}\otimes\hat{b}) : \epsilon^{\times} \\ &= \frac{1}{2}(1+\cos^2\theta)\cos(2\phi)\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi \end{aligned}$$

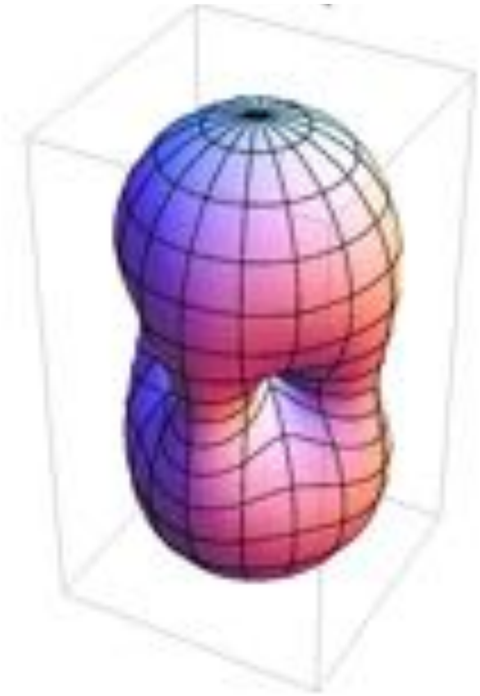
Antenna Pattern Functions

F_+



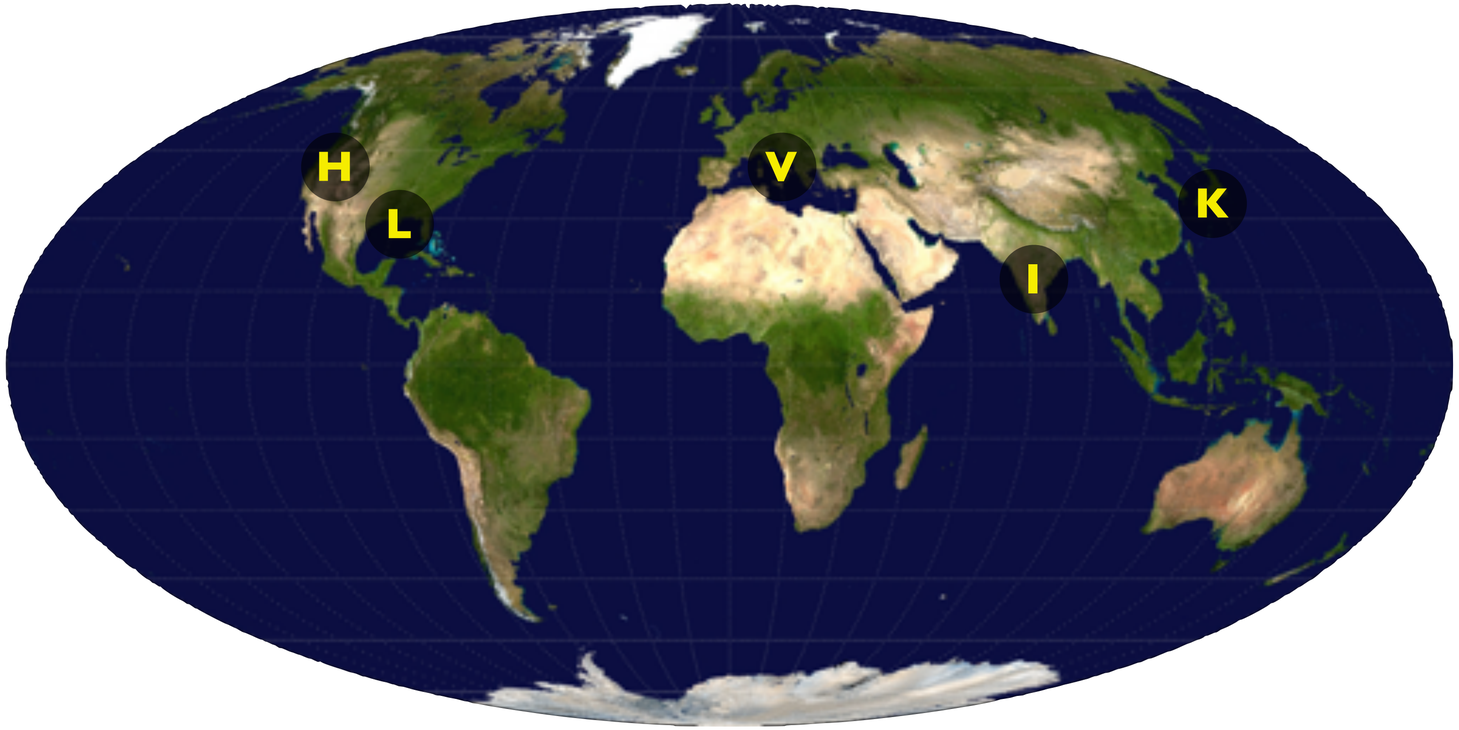
$F_\times$

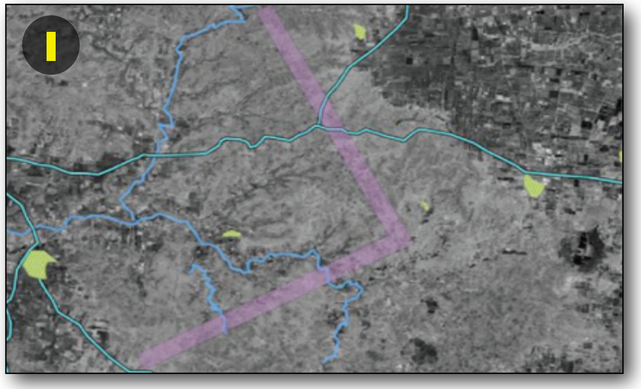
$F = \sqrt{F_+^2 + F_\times^2}$

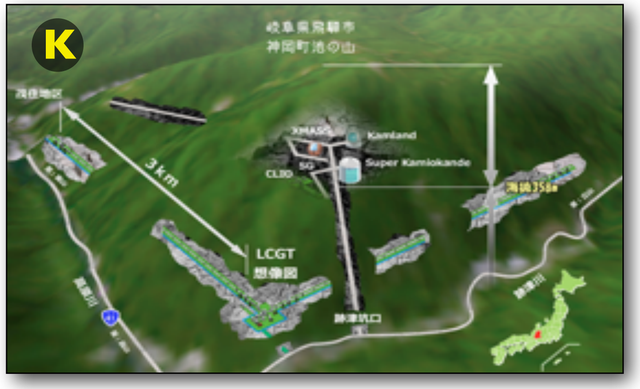


Polarization averaged

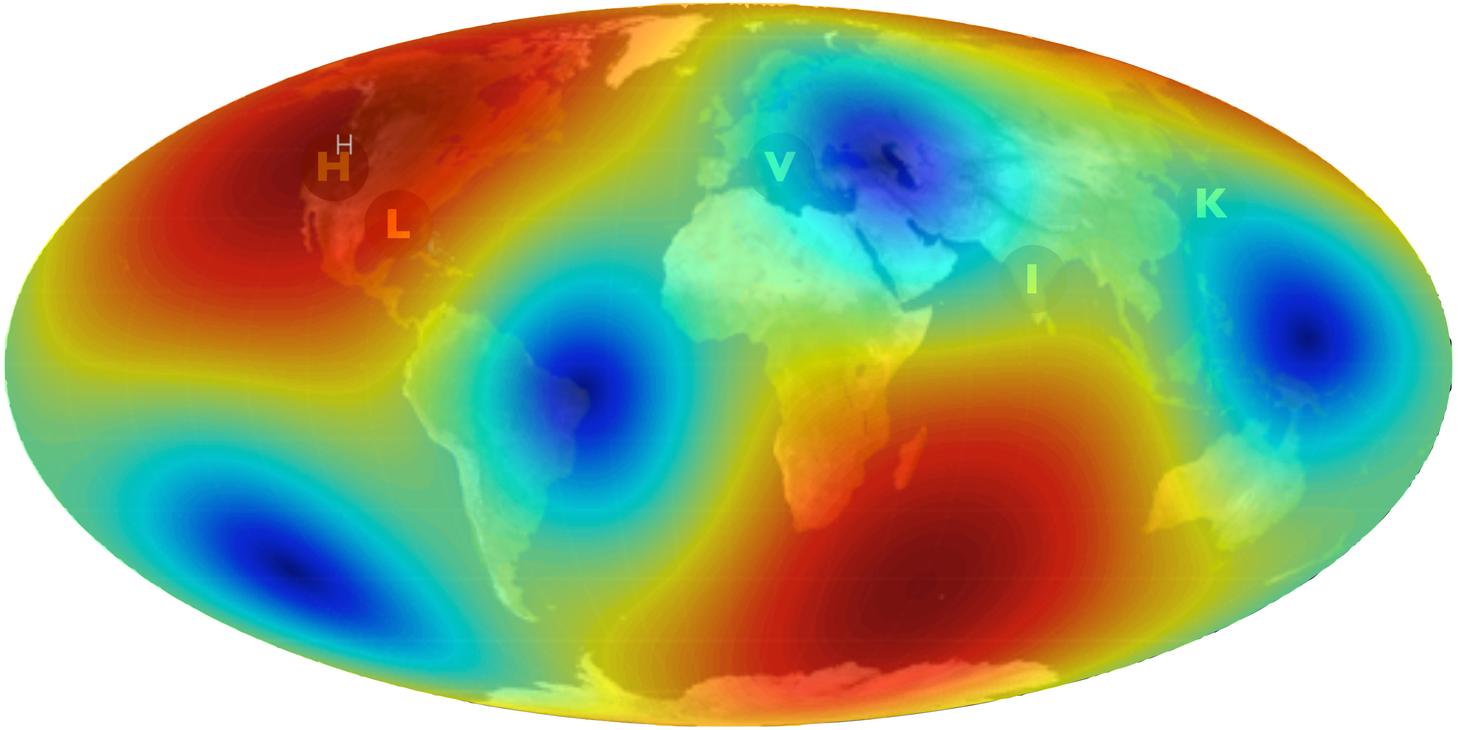
Terrestrial Network

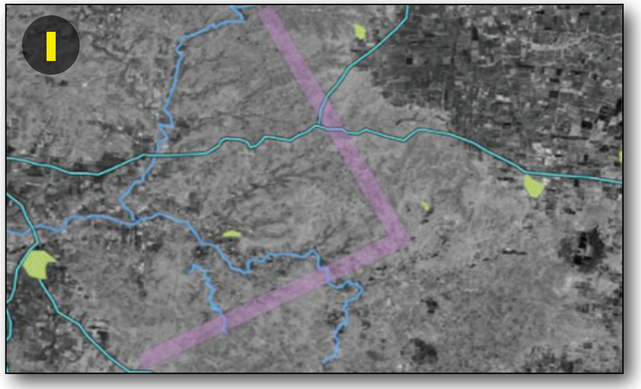


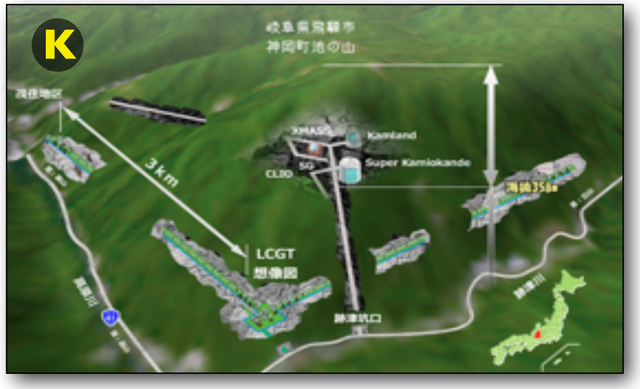




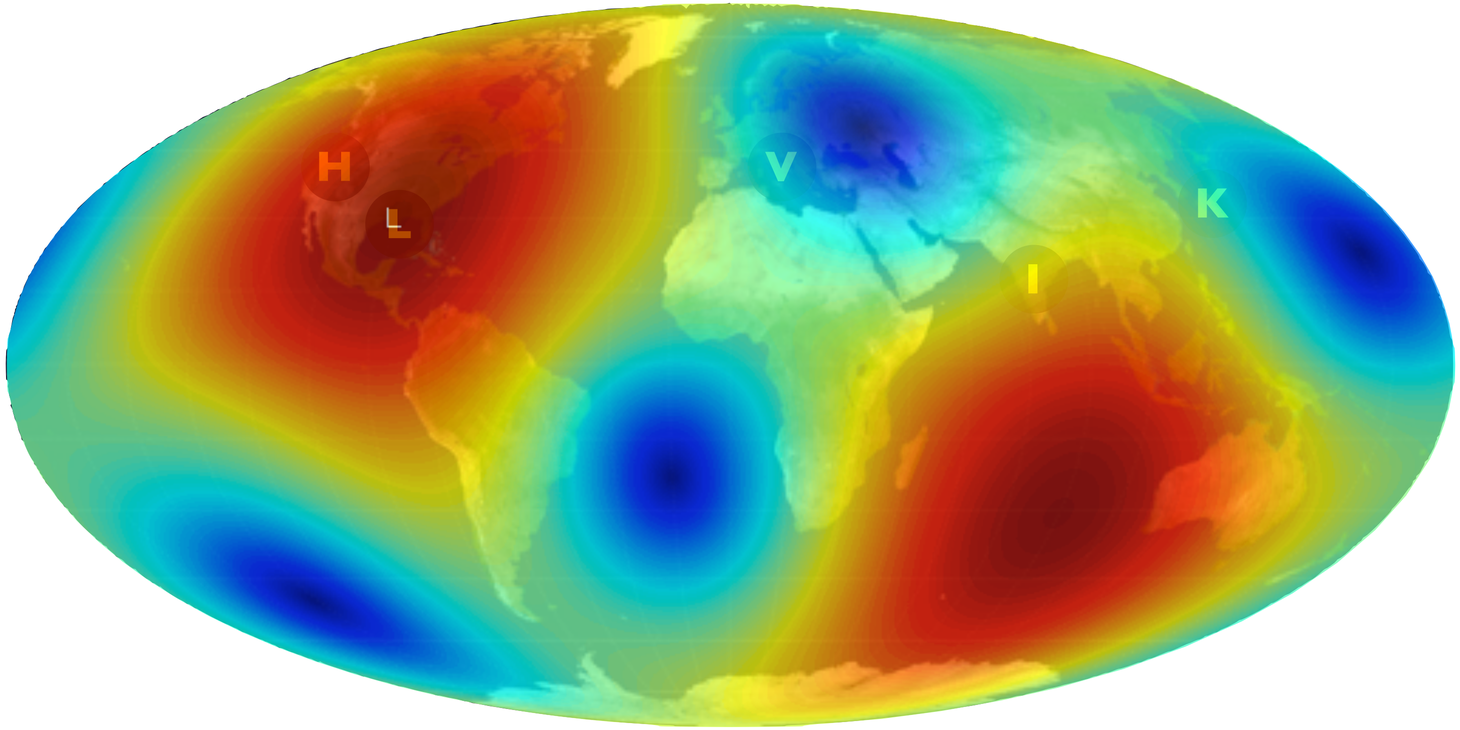
Terrestrial Network

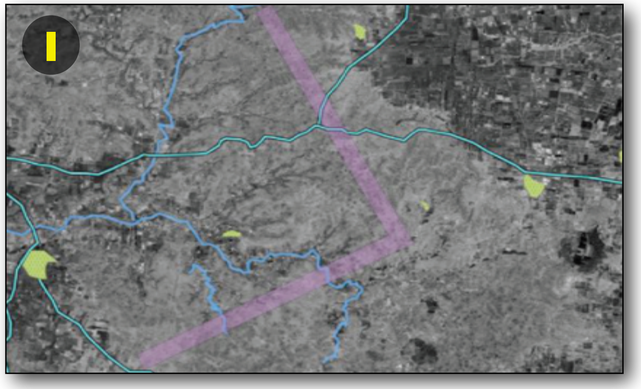


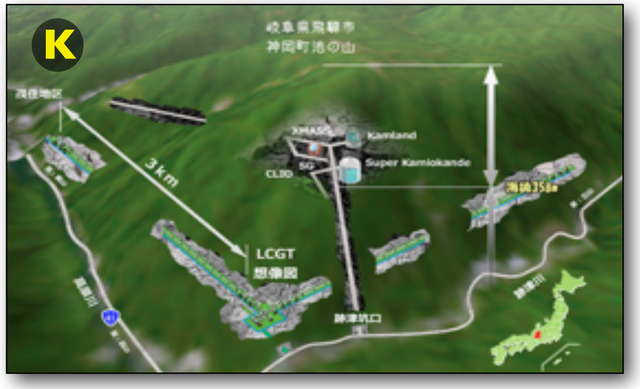




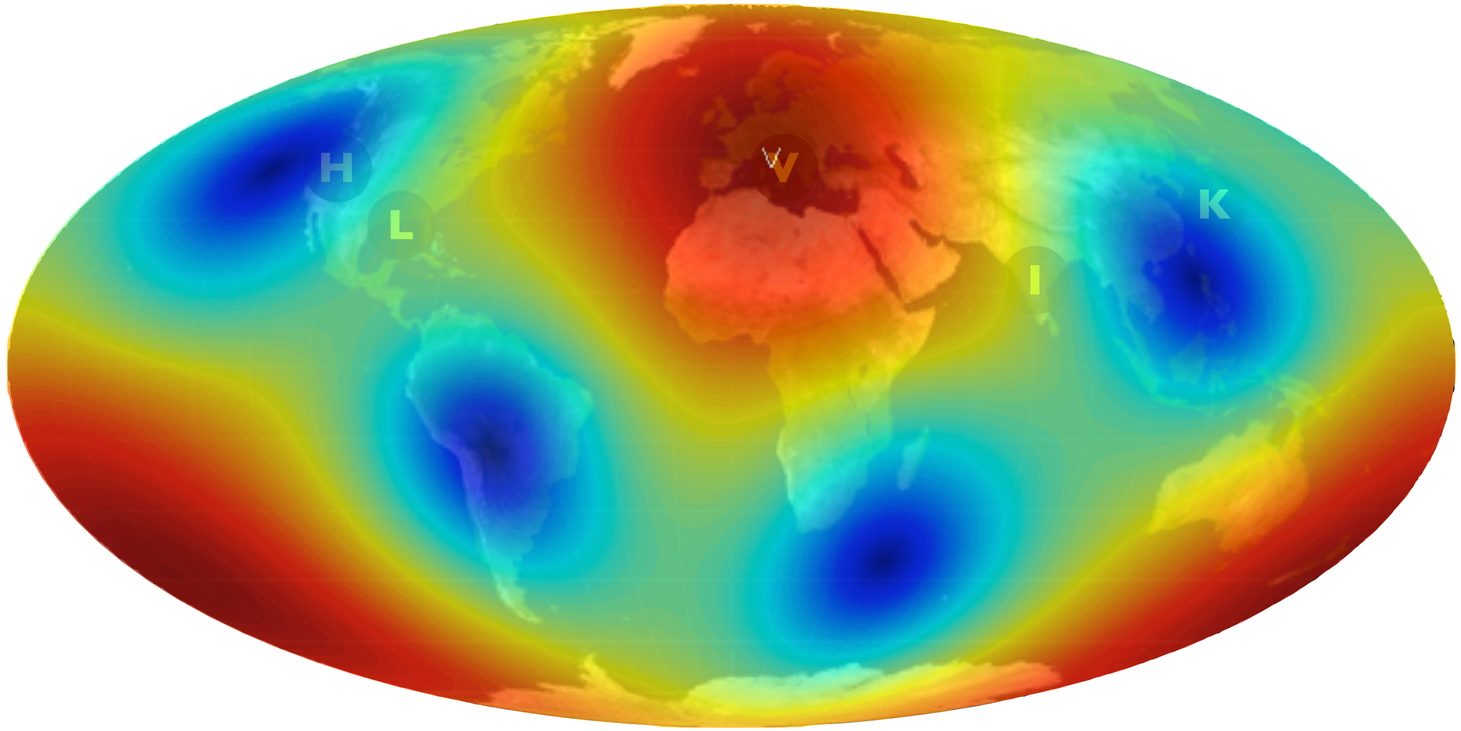
Terrestrial Network

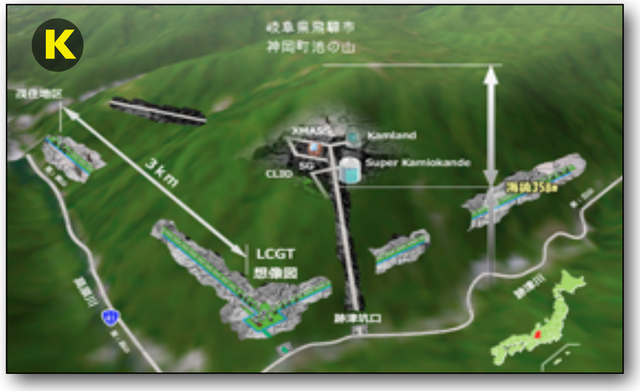




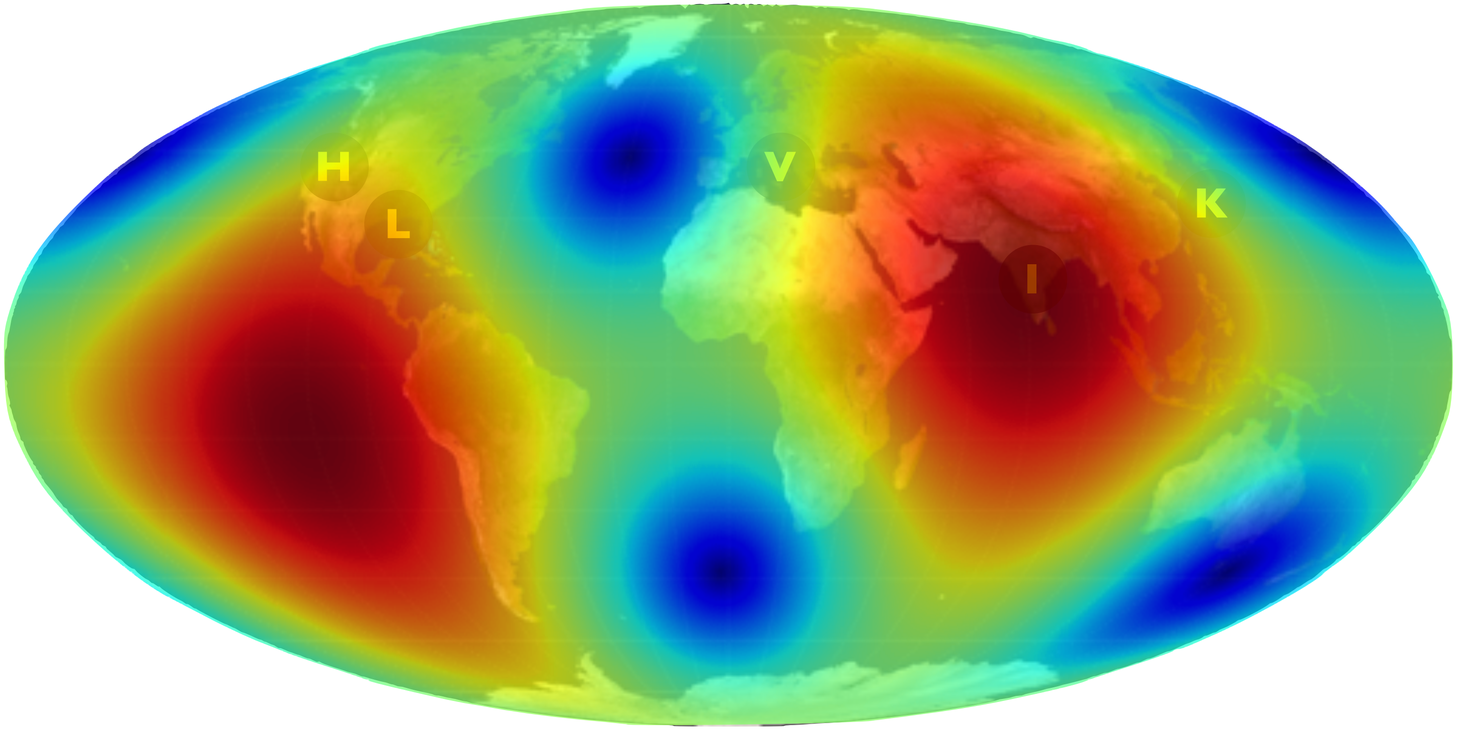


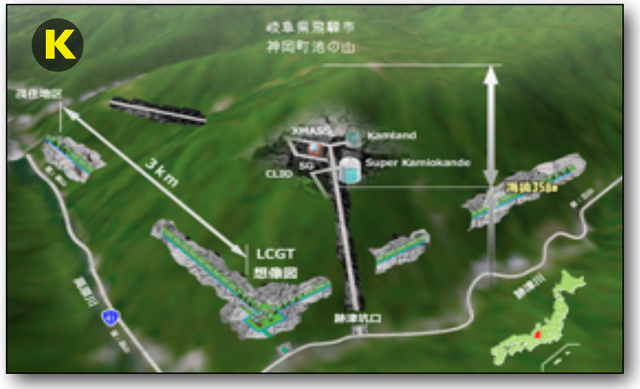
Terrestrial Network



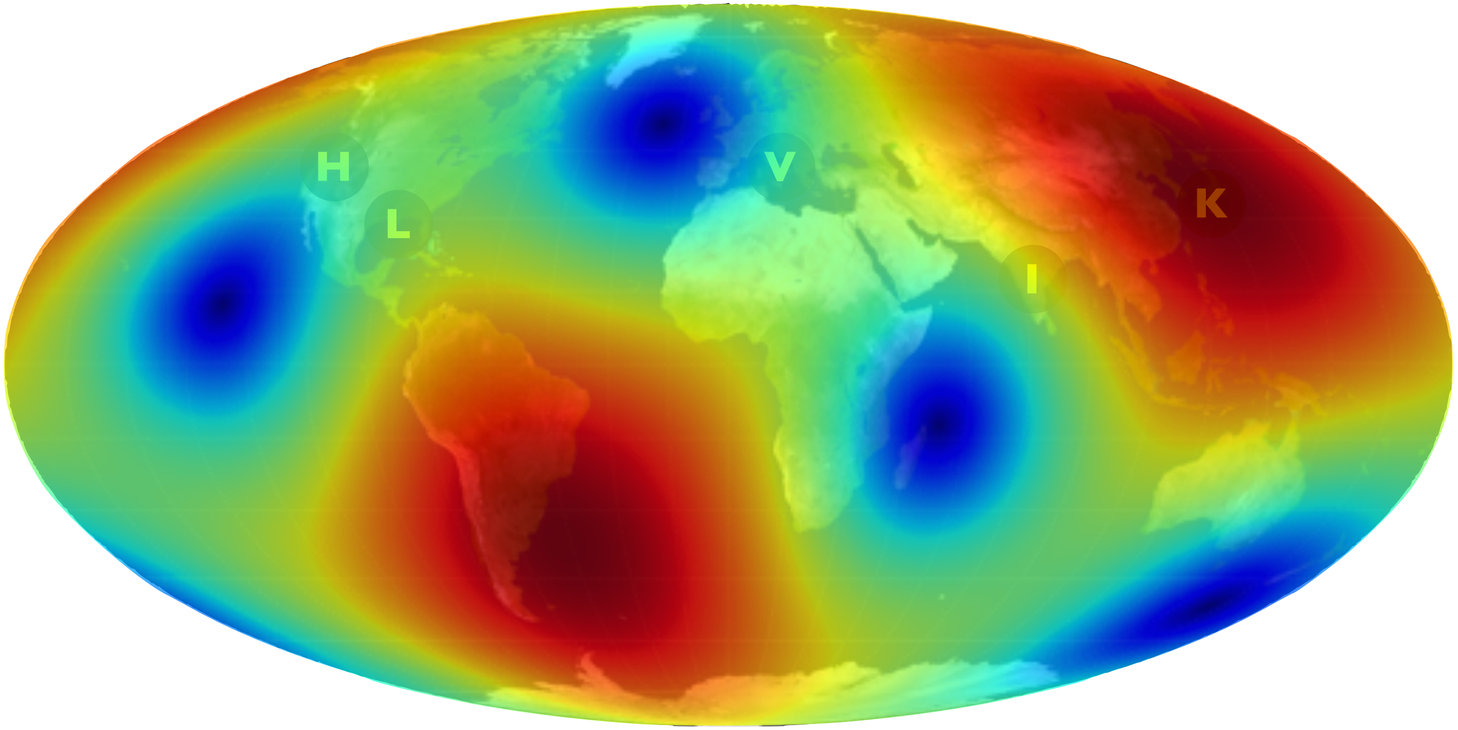


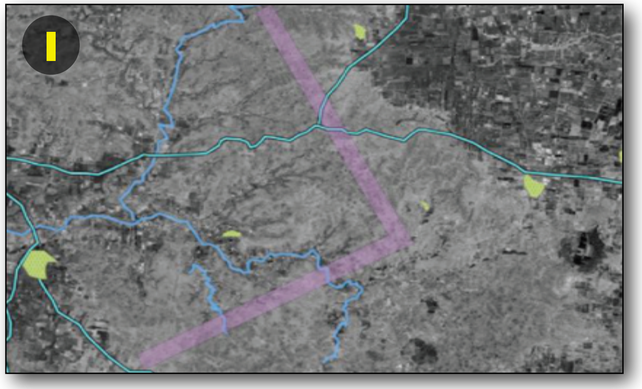
Terrestrial Network



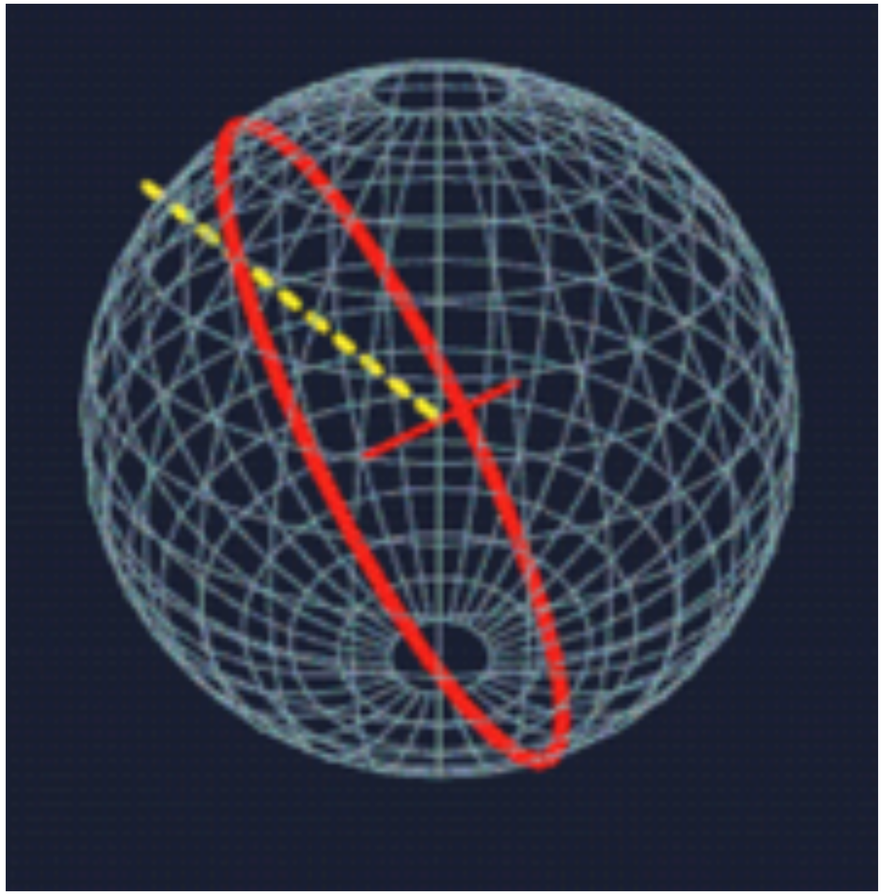


Terrestrial Network

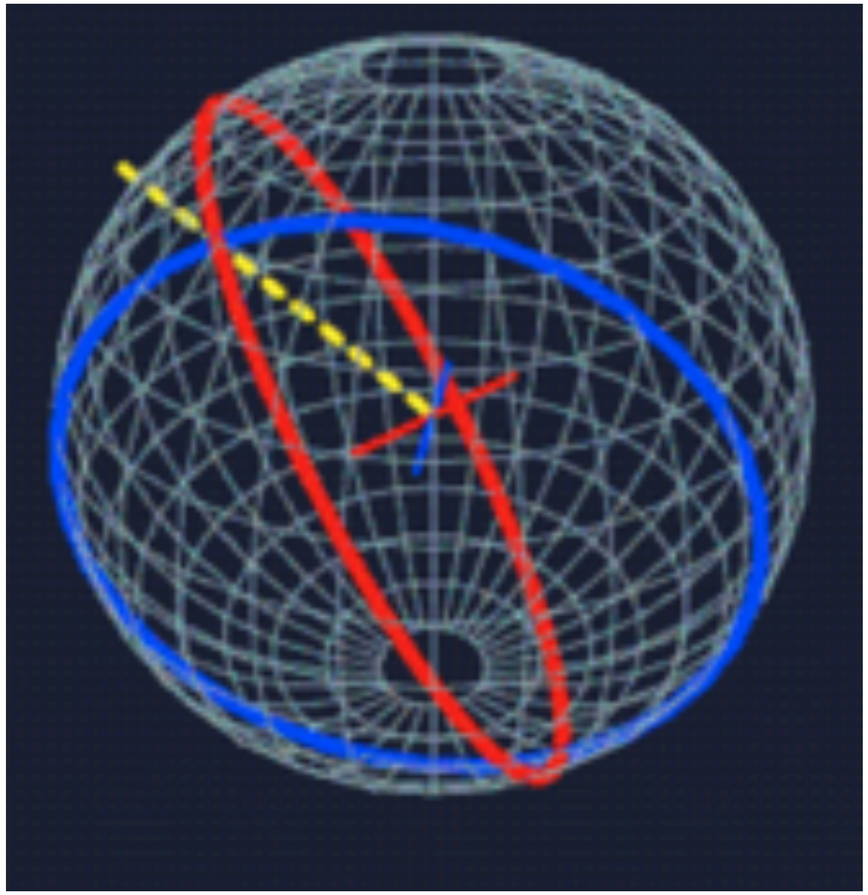




Time of Arrival Triangulation

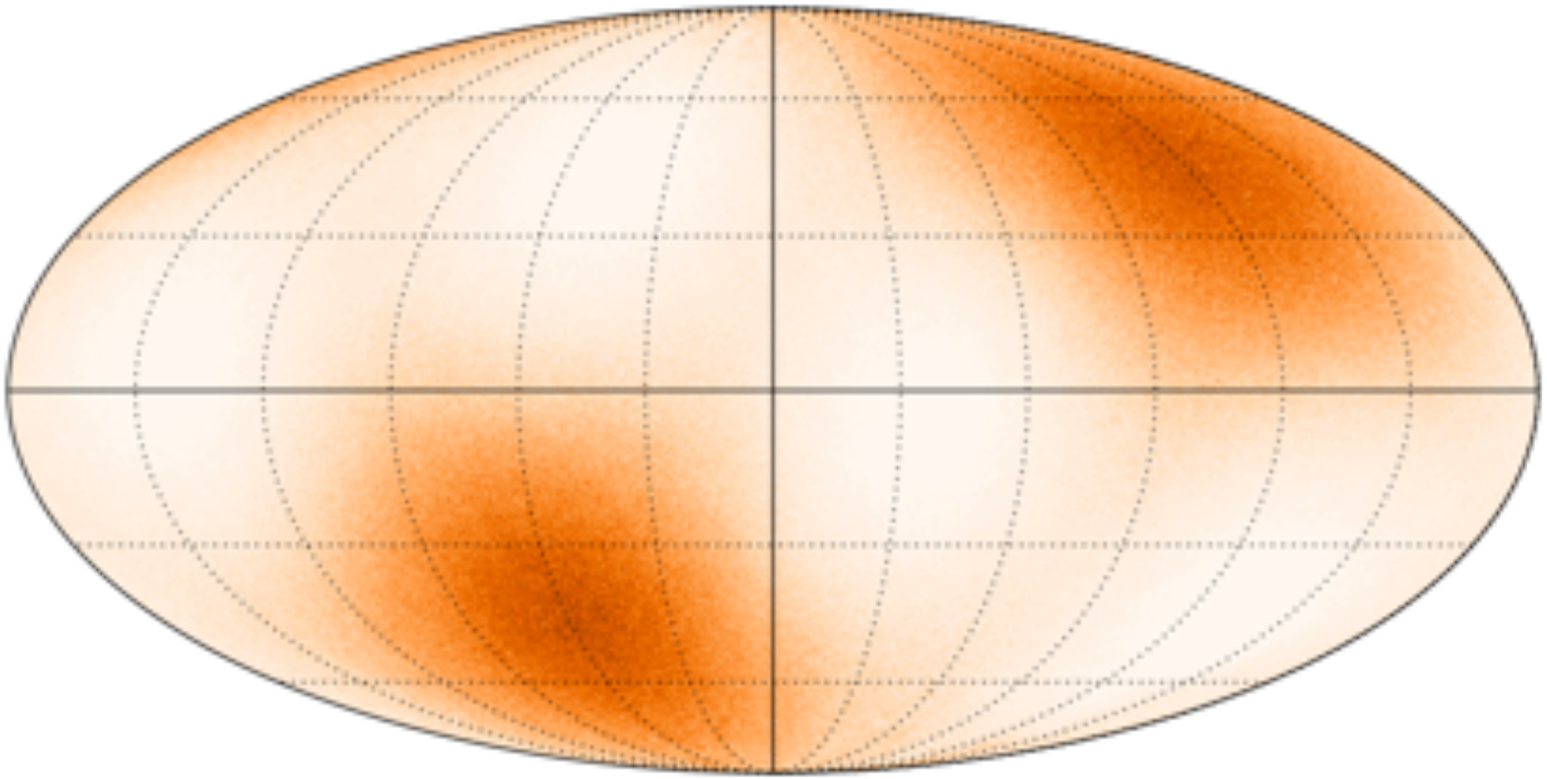


LIGO Hanford + LIGO Livingston



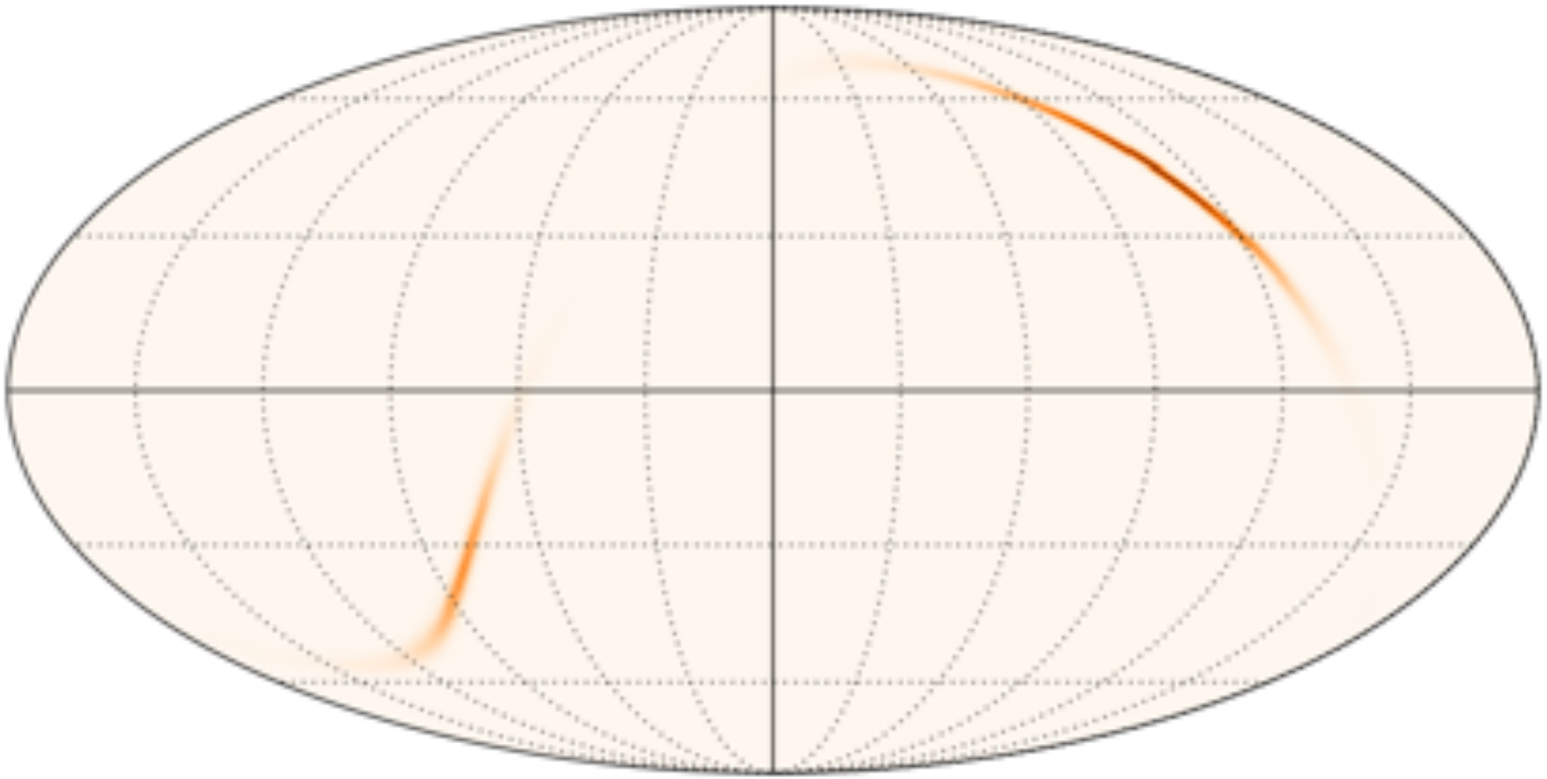
LIGO Hanford + LIGO Livingston + Virgo

Triangulating the Source



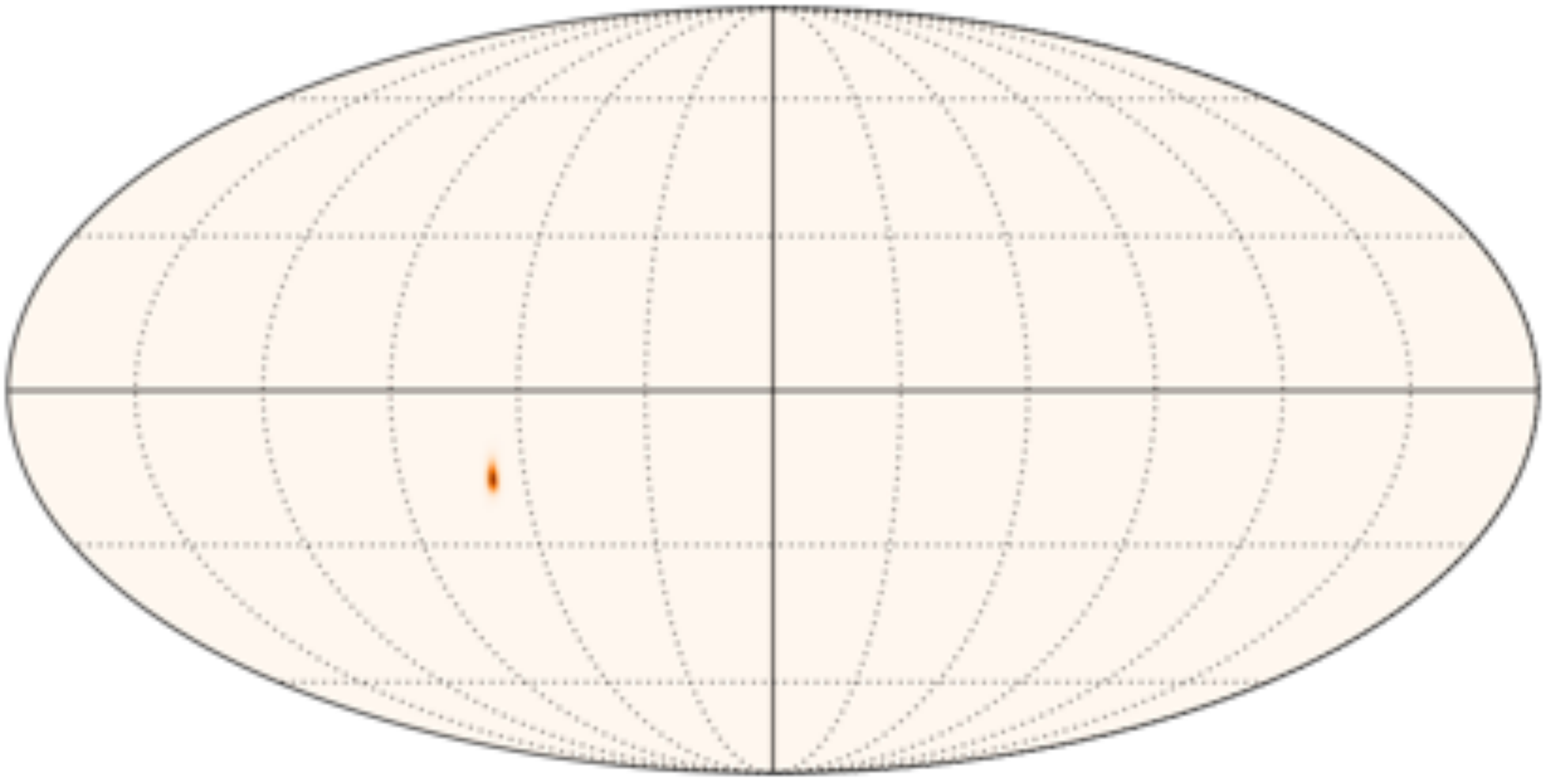
Hanford

Triangulating the Source



Hanford + Livingston

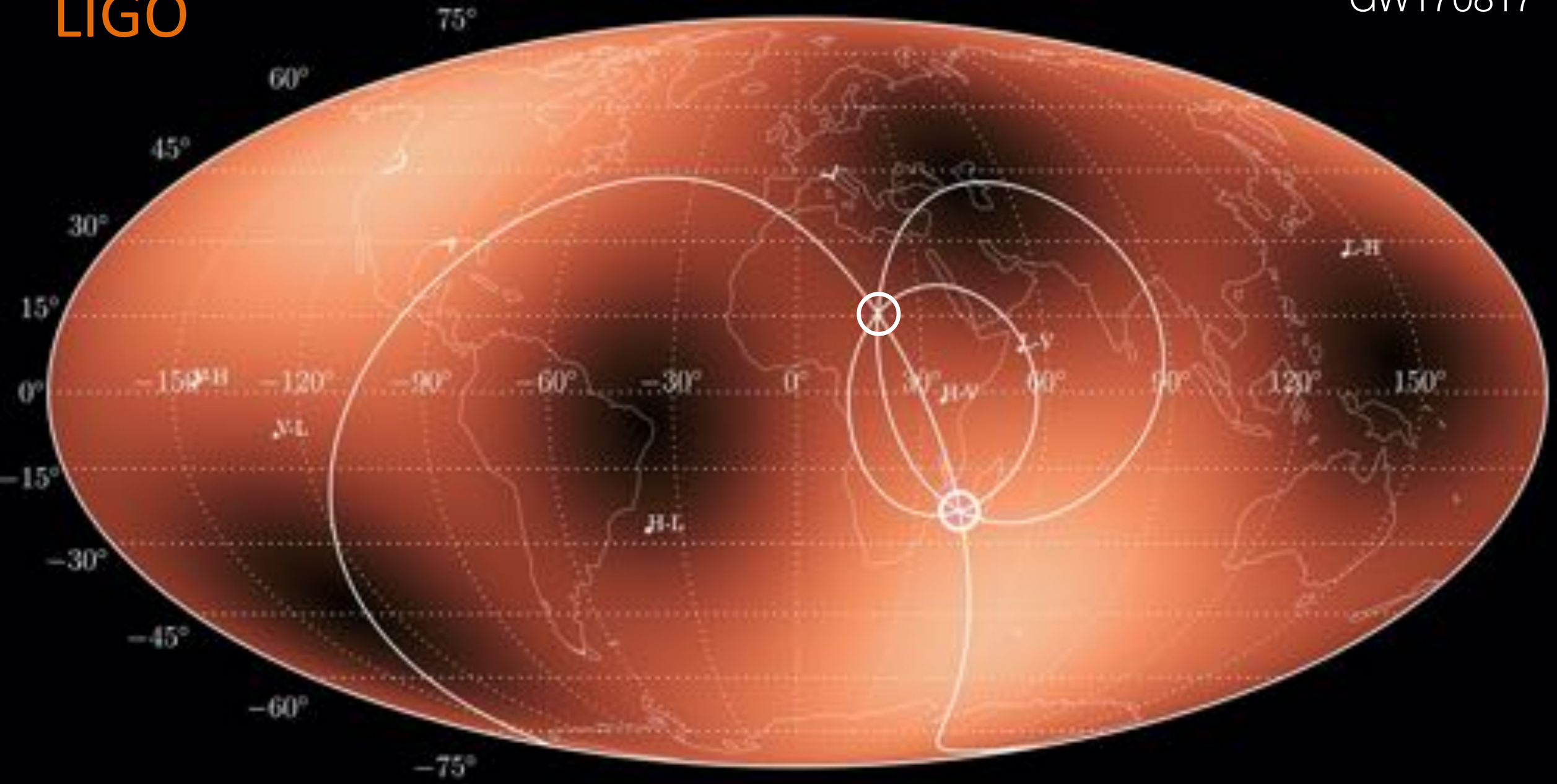
Triangulating the Source



Hanford + Livingston + Virgo

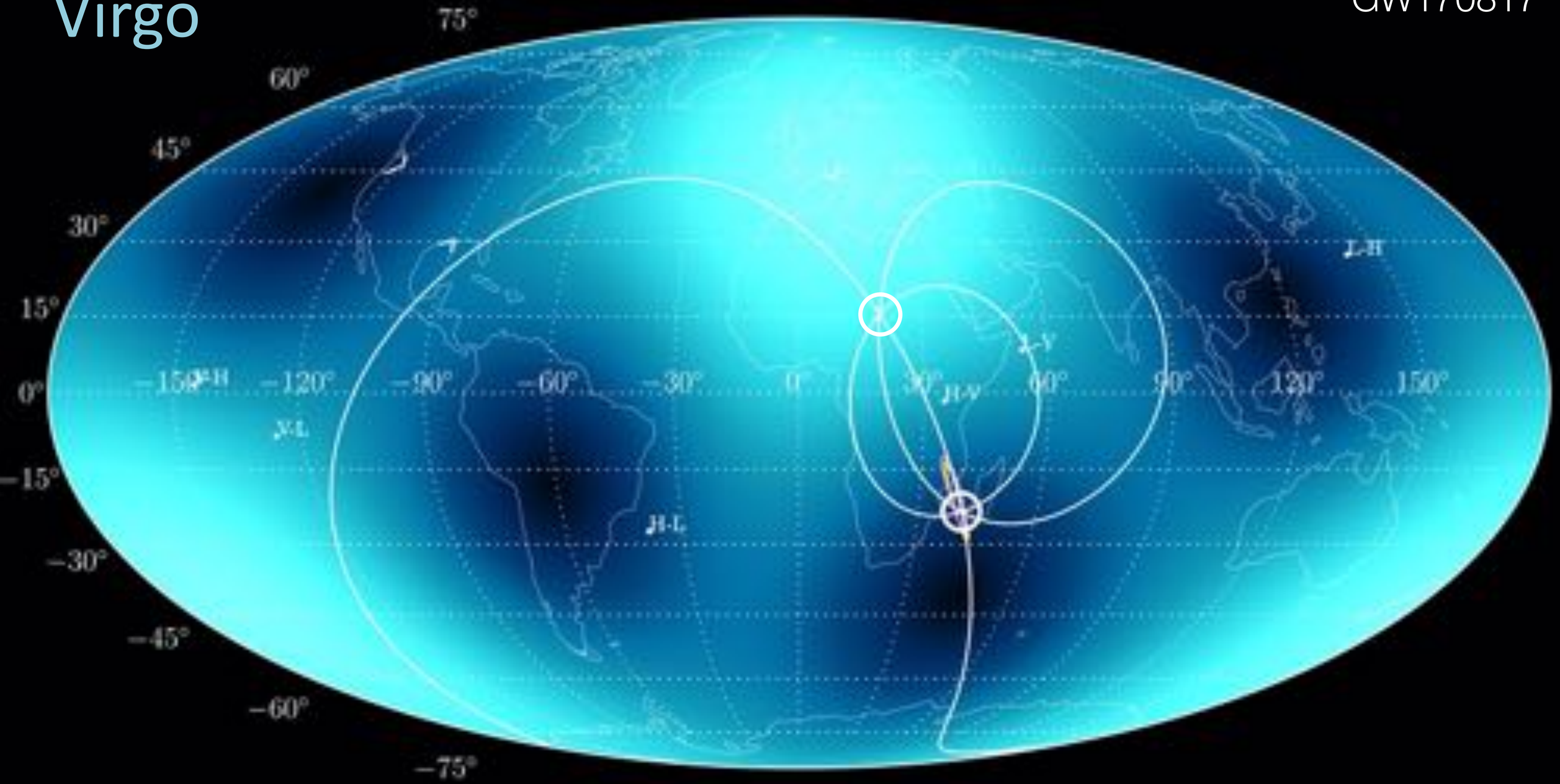
LIGO

GW170817



Virgo

GW170817



Beyond the low frequency approximation

$$\Delta\tau_{12} = \frac{(\hat{a}\otimes\hat{a}) : \mathbf{H}[u_1, u_2]}{2(1-\hat{k}\cdot\hat{a})} \qquad \mathbf{H}[u_1, u_2] = \int_{u_1}^{u_2} \mathbf{h}(u)\, du \qquad (u = k_\alpha x^\alpha)$$

Example: $$\mathbf{h}(u) = A\cos(\omega(t - \hat{k}\cdot\mathbf{x}))\,\boldsymbol{\epsilon}^+$$

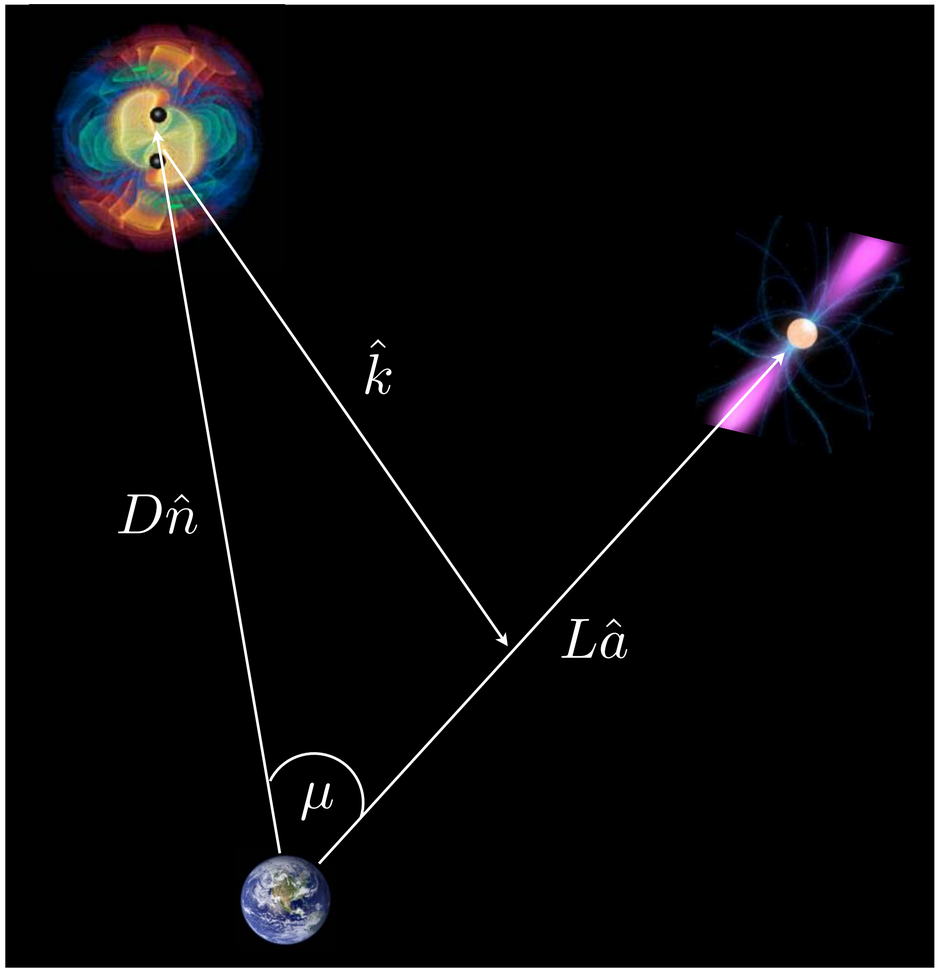
$$\Delta\tau_{12} = \underbrace{\frac{L}{2}((\hat{a}\otimes\hat{a}) : \boldsymbol{\epsilon}^+)}_{\text{Long wavelength one-arm antenna pattern}}\,\underbrace{\mathrm{sinc}\left[\frac{\omega L}{2}(1-\hat{k}\cdot\hat{a})\right]}_{\text{Finite arm-length correction to antenna pattern}}\cos\left[\underbrace{\omega(t + \frac{L}{2} - \frac{\hat{k}\cdot(\mathbf{x_1}+\mathbf{x_2})}{2}}_{\text{Phase of the wave at midpoint of arm}}\right]$$

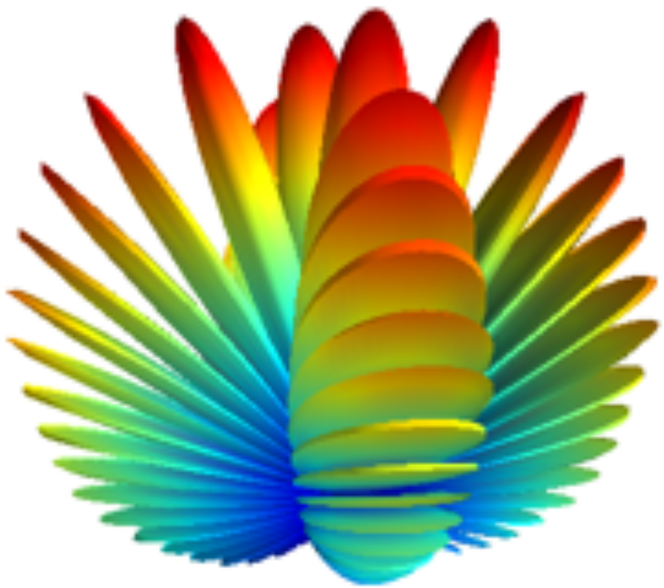
Pulsar Timing

$$\tau_{\rm GW}(t) = -\frac{L}{2}\int_{-1}^{0} (\hat{a}\otimes\hat{a}) : \mathbf{h}(t+L\xi, -\hat{a}\xi L)\, d\xi$$

(note that here the photons propagate in the $-\hat{a}$ direction)

$$\hat{k} = -\hat{n} - \xi\frac{L}{D}(\hat{a} - \hat{n}\cos\mu)$$





GW

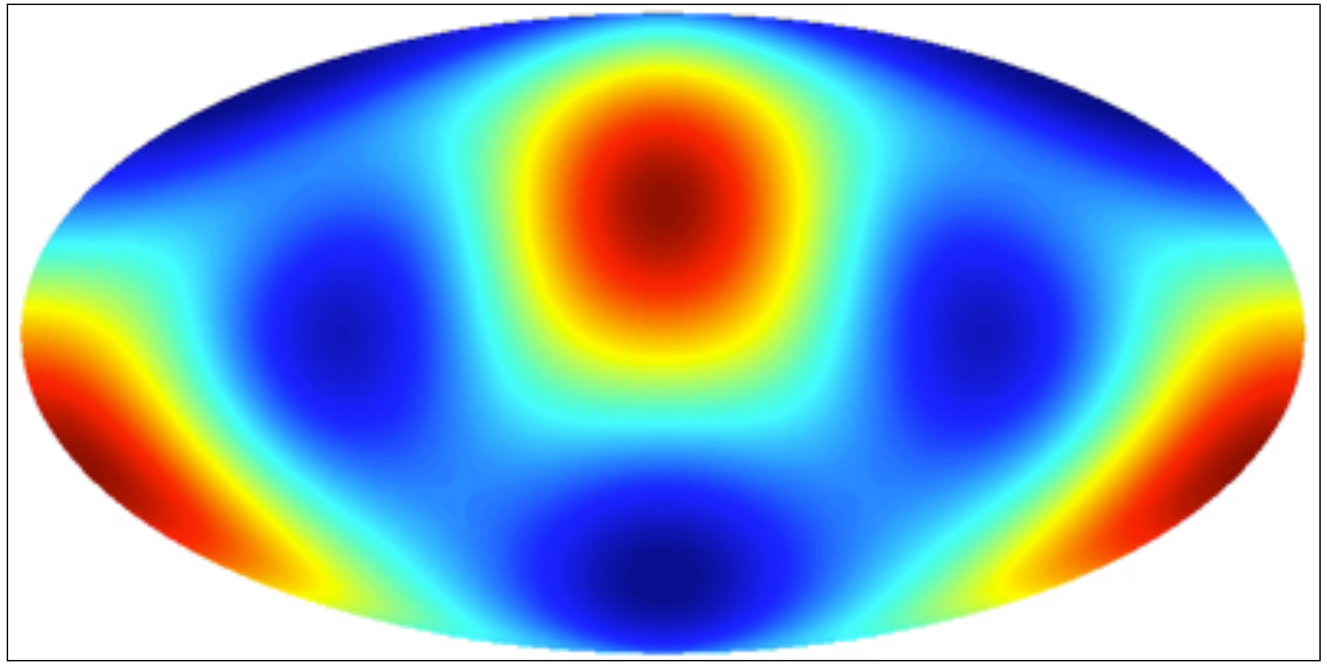
$$fL = 10,\ \psi = 0$$

$$(\hat{a}\otimes\hat{a}) : \mathbf{H} = (1+\cos\mu)(H_+\cos 2\psi + H_\times \sin 2\Psi)$$

(Ignoring *L*/*D* amplitude corrections)

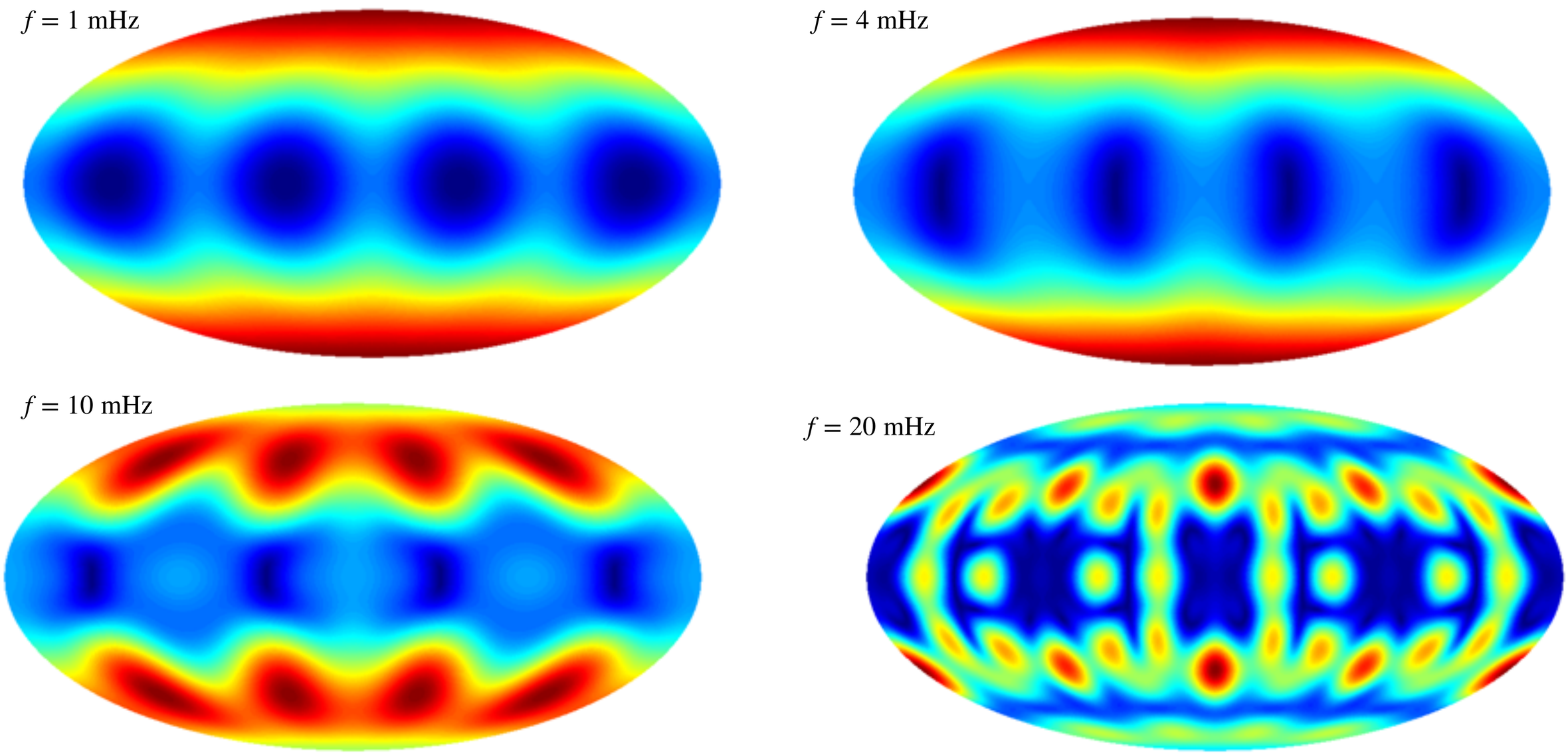
Laser Interferometer Space Antenna

Source localization via amplitude and frequency modulation



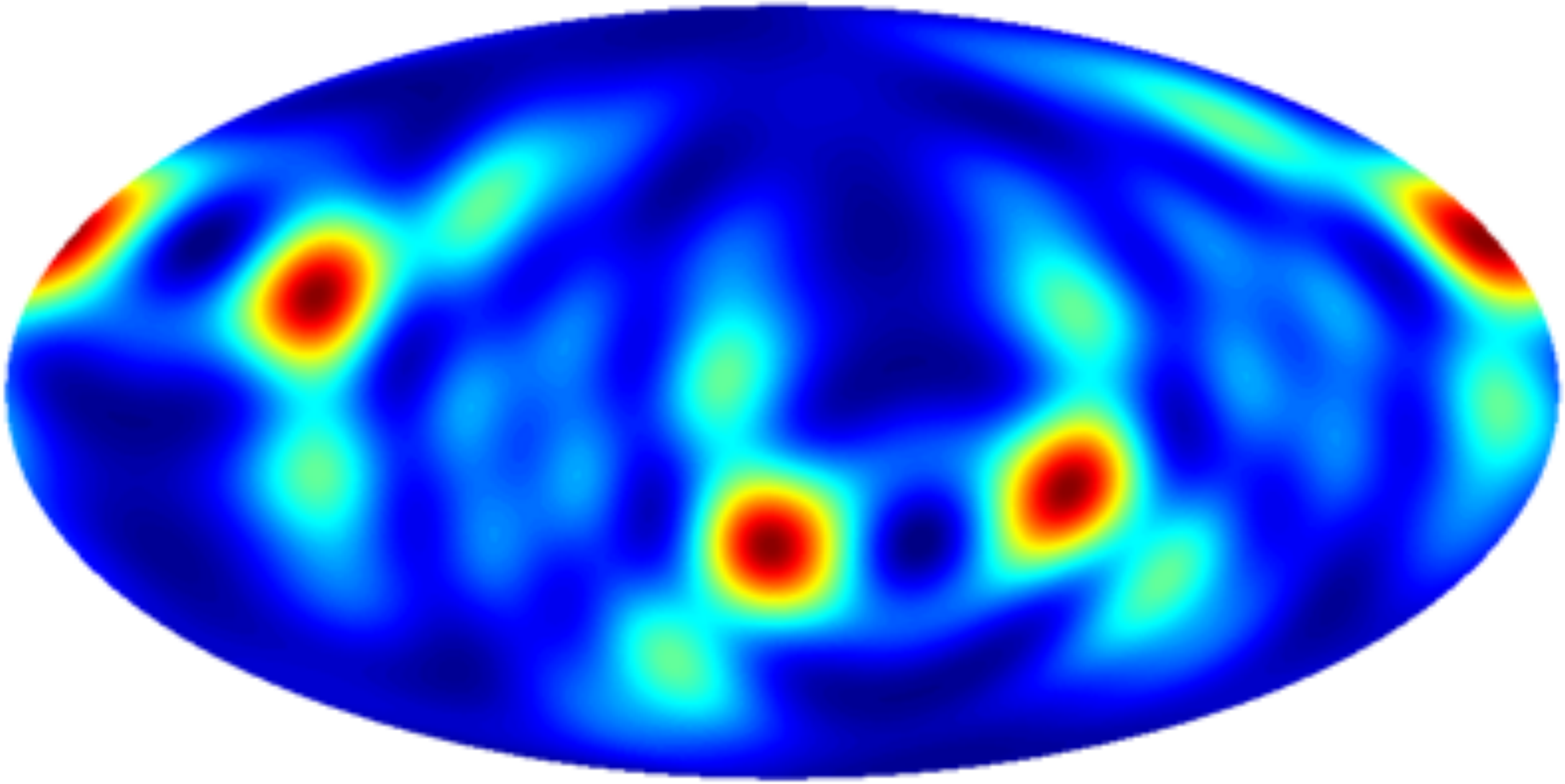
Low frequency response

LISA Antenna Patterns (Detector Frame)



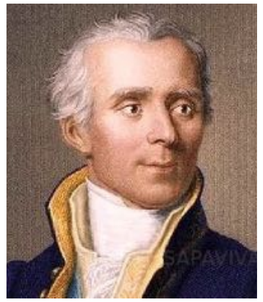
N. J. Cornish, Class.Quant.Grav.18:4277, (2001)

LISA 20 mHz



Data Analysis 101 - Laplace & Gauss

Pierre Simon de Laplace

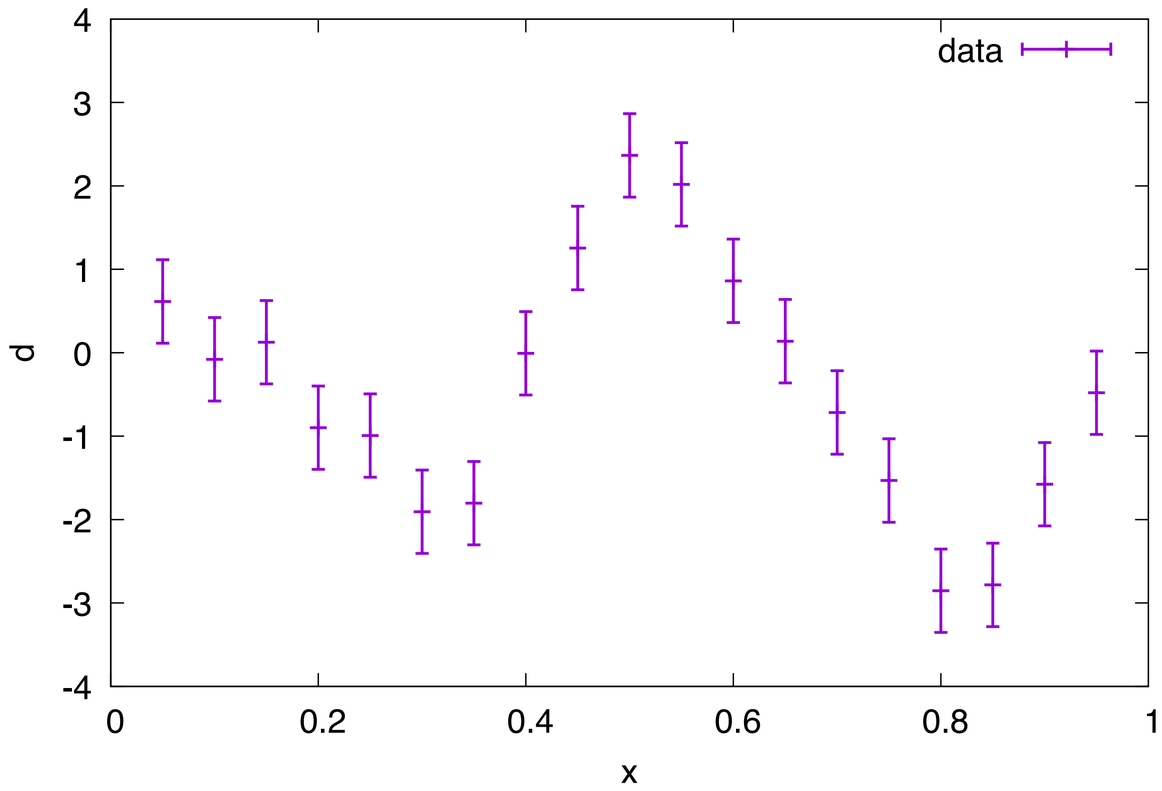


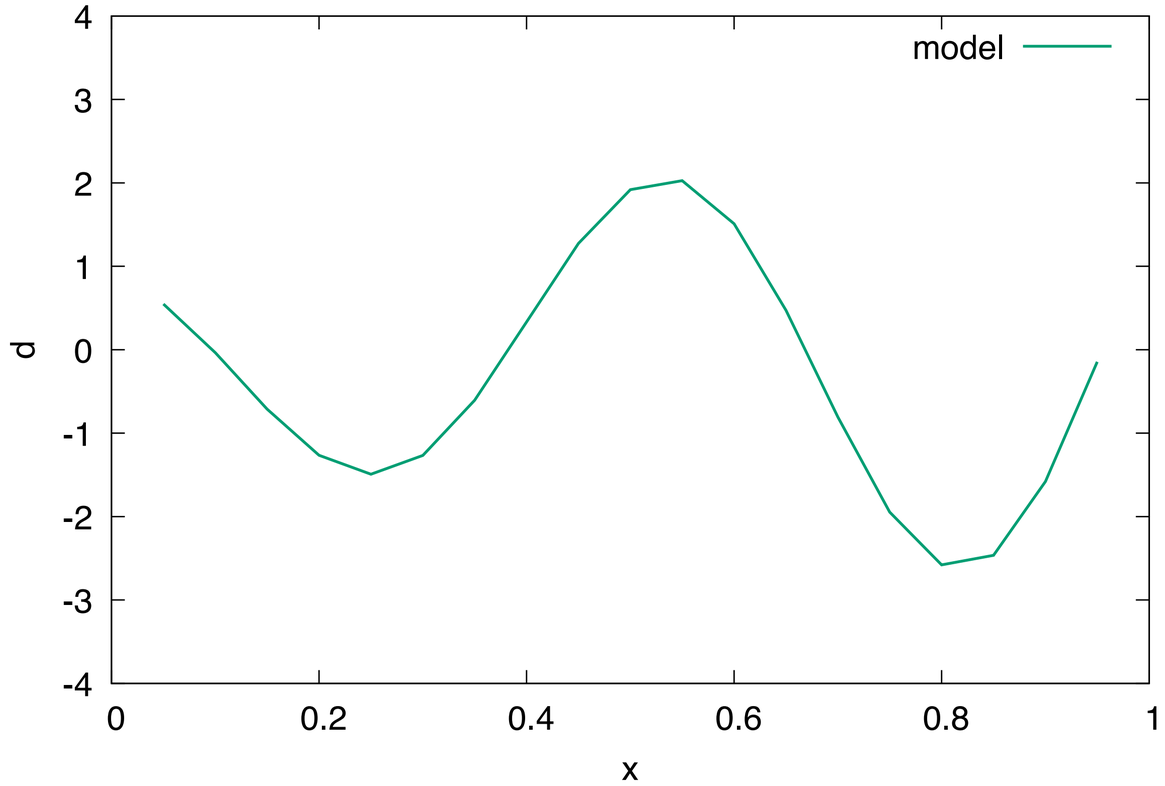
Memoir on the Probability of Causes of Events (1774)
Analytical Theory of Probability (1812)

Carl Friedrich Gauss

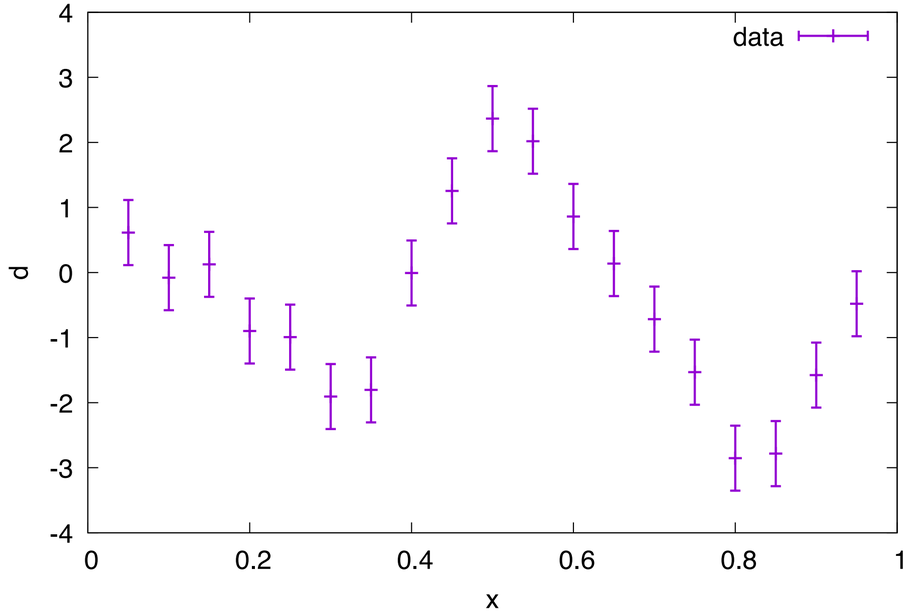
Z. Astronom. Verwandte Wiss. 1 185, (1816)

Laplace developed Bayesian Inference. Gauss developed maximum likelihood estimation.
Gauss introduced the normal distribution, Laplace explained its ubiquity (CLT).

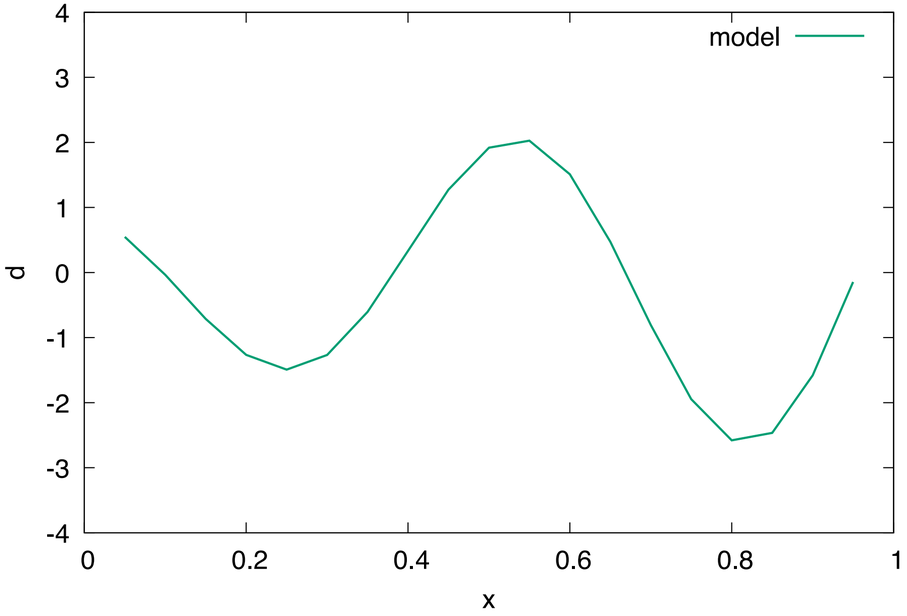




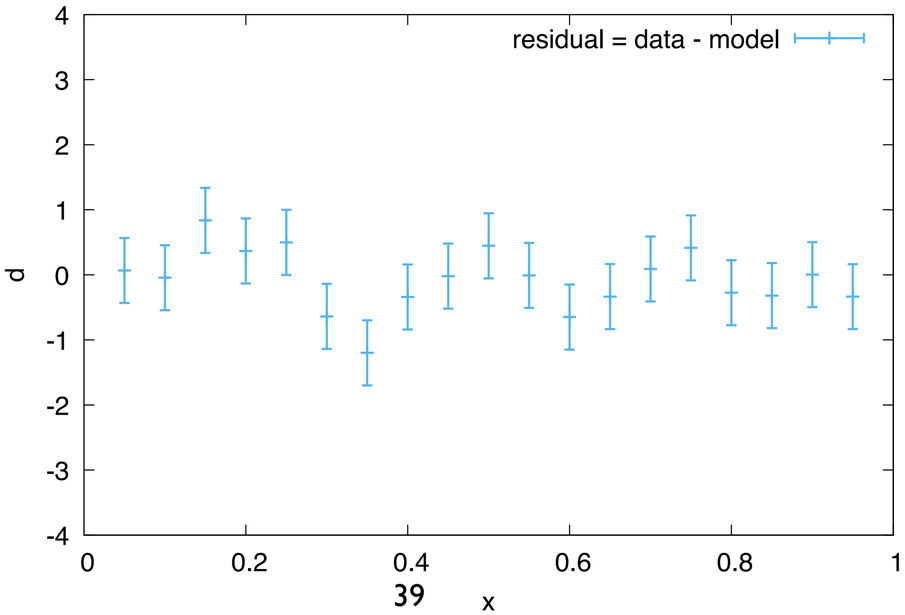
Data Analysis 101



−



=



Data Analysis 101

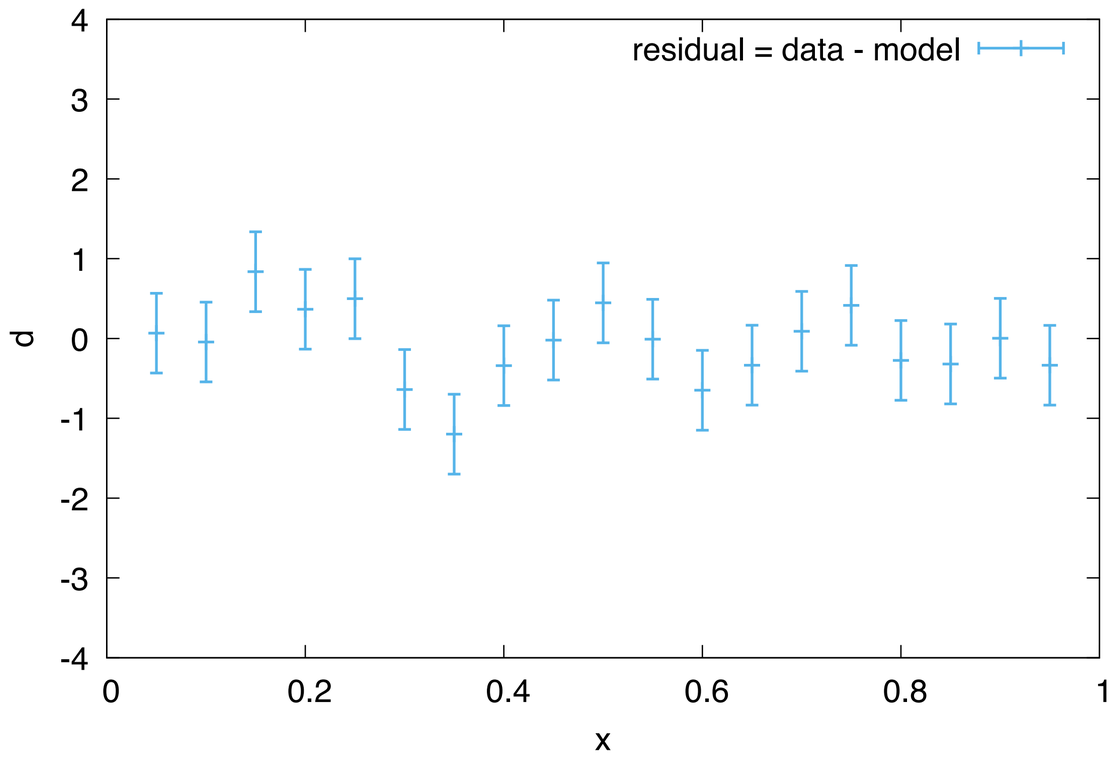
$$d = h + n \quad \Rightarrow \quad d - h = n$$

The residuals should follow the noise distribution

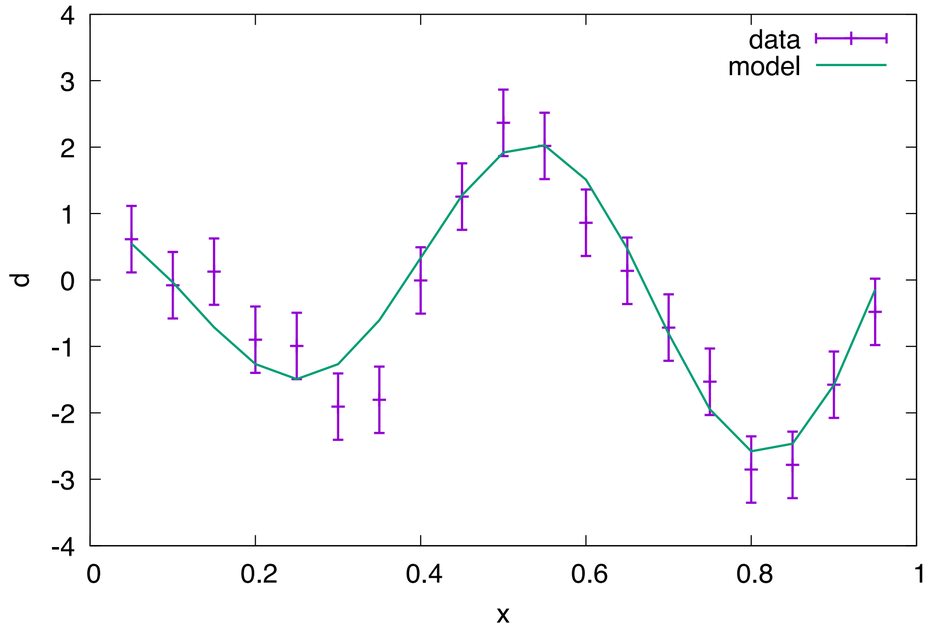
In this example the noise was uncorrelated between samples and draw from a Gaussian distribution with a fixed standard deviation ("stationary white noise")

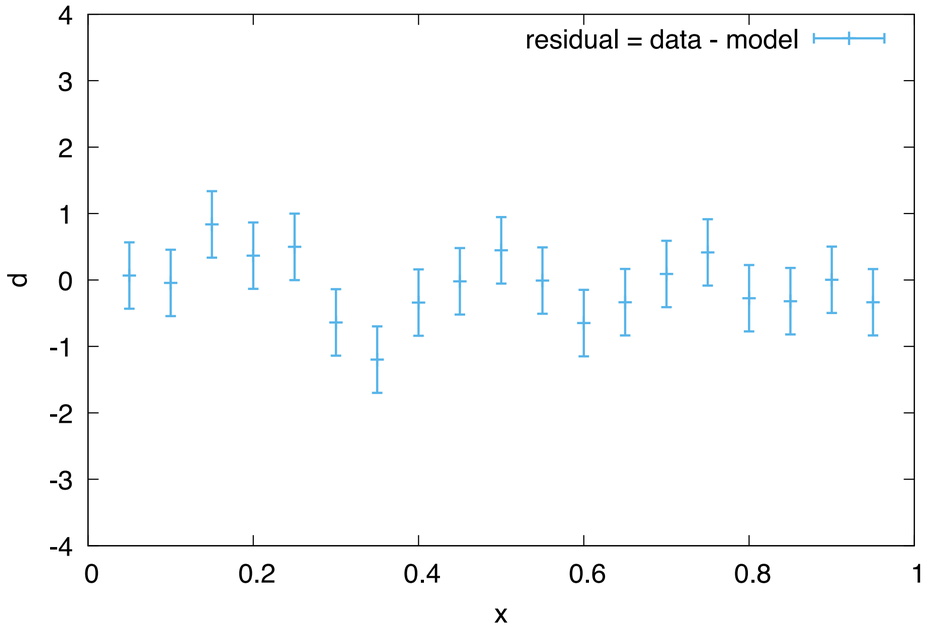
$$p(n_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n_i^2}{2\sigma^2}}$$

$$p(n) = \prod_i p(n_i) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\sum_{i=1}^{N} \frac{n_i^2}{2\sigma^2}}$$



Data Analysis 101





The likelihood of observing the data d given the model h is simply

$$p(d-h) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\sum_{i=1}^{N} \frac{(d_i - h_i)^2}{2\sigma^2}}$$

Data Analysis 101

If the noise is correlated between data samples ("colored noise"), and/or if the amplitude of the noise changes from sample to sample ("heteroskedastic" aka "non-stationary"), then we need to generalize the Gaussian likelihood:

$$p(d-h) = \frac{1}{(\det(2\pi\mathbf{C}))^{N/2}}\, e^{-\frac{1}{2}(d_i - h_i)C_{ij}^{-1}(d_j - h_j)}$$

In the previous example the noise correlation matrix was promotional to the identity matrix: $C_{ij}^{\mathrm{SWN}} = \delta_{ij}\,\sigma^2$

The quantity in the exponent is the chi squared goodness of fit

$$\chi^2 = (d_i - h_i)C_{ij}^{-1}(d_j - h_j) \equiv (d-h\,|\,d-h)$$

Here I have introduced the noise weighted inner product

$$(a\,|\,b) = a_i\, C_{ij}^{-1}\, b_j$$

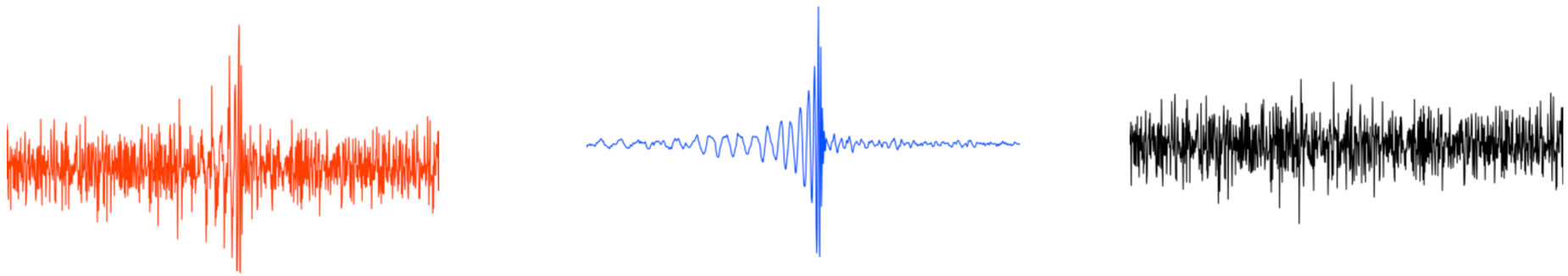
Data Analysis 101

Gravitational wave data comes in the form of a time series. Computing the noise-weighted inner product in the time domain is costly since the matrix is not diagonal. If the noise is *stationary,* i.e. has statistical properties that are unchain with time, then the noise correlation matrix is diagonal in the frequency domain. That is why most GW analysis is done in the frequency domain.

$$(a\,|\,b) = a_i\, C_{ij}^{-1}\, b_j = 2\sum_f \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)}$$

The factor of 2 is because we only sum over positive frequencies. The quantity $S_n(f)$ is the one-sided noise spectral density (PSD)

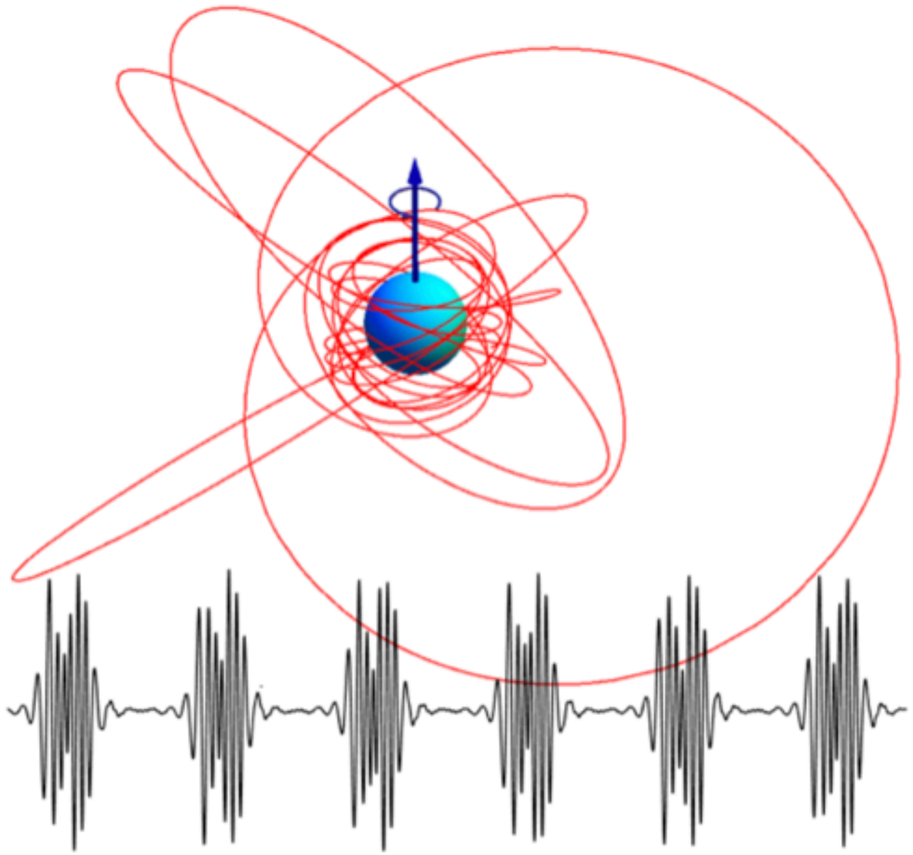
We often talk about “whitened” data or waveforms. This is simply $\tilde{a}^W(f) = \tilde{a}(f)/S_n(f)^{1/2}$. Can transform this back to the time domain:

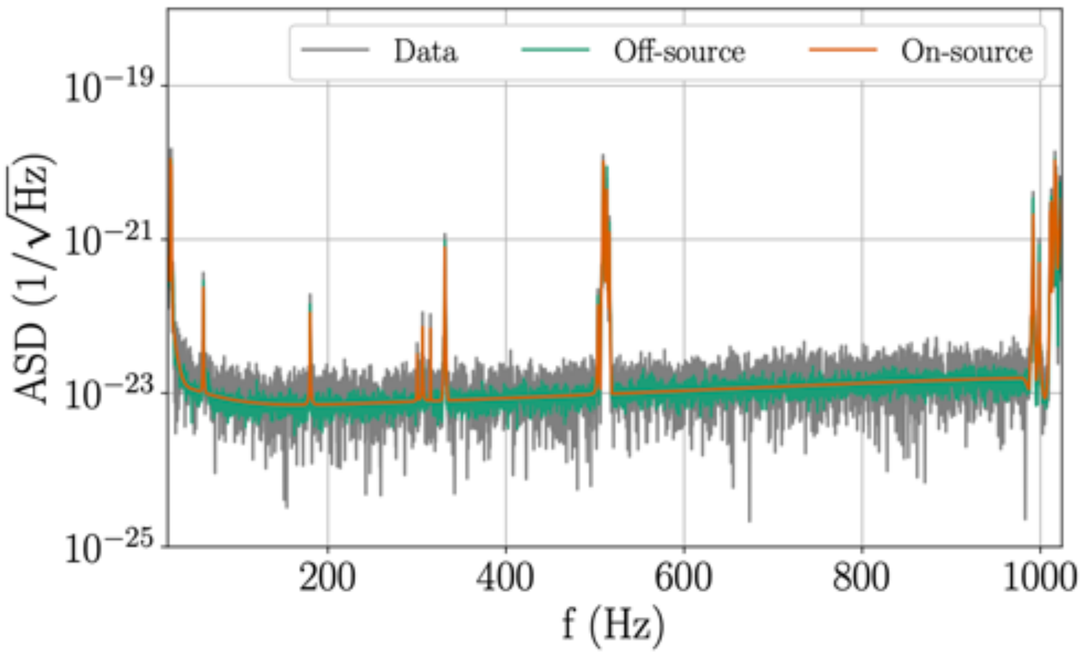


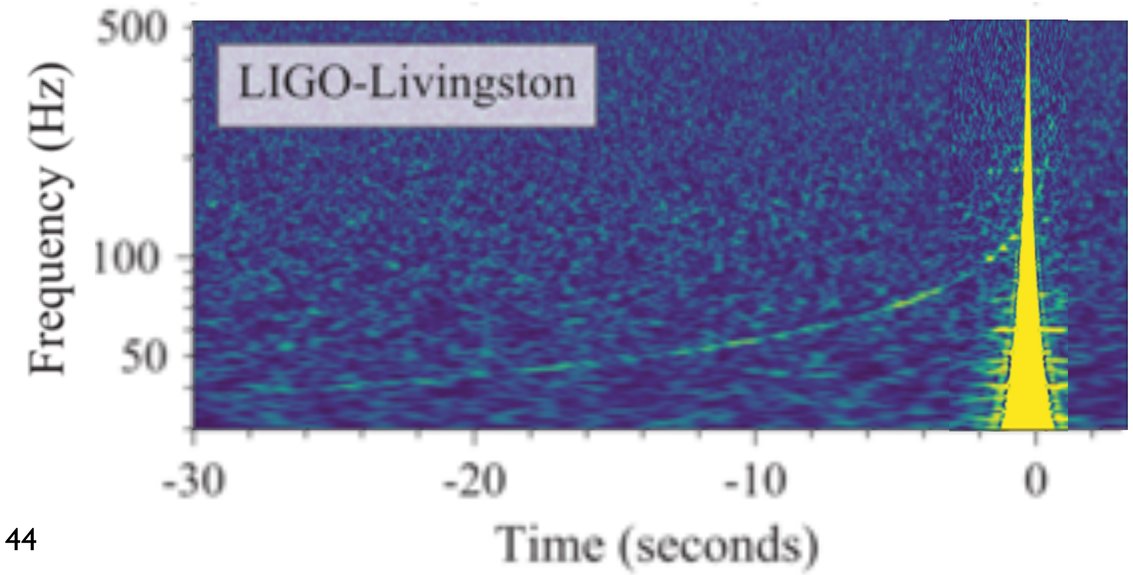
data - signal = noise

Challenges for Gravitational Wave Data Analysis

- Complicated waveforms, as many as 17 parameters
- Noise properties have to be estimated along with the signals
- Non-Gaussian noise transients have to be modeled/mitigated







The Bayesian Way

Likelihood
(noise model)

Prior
(signal model)

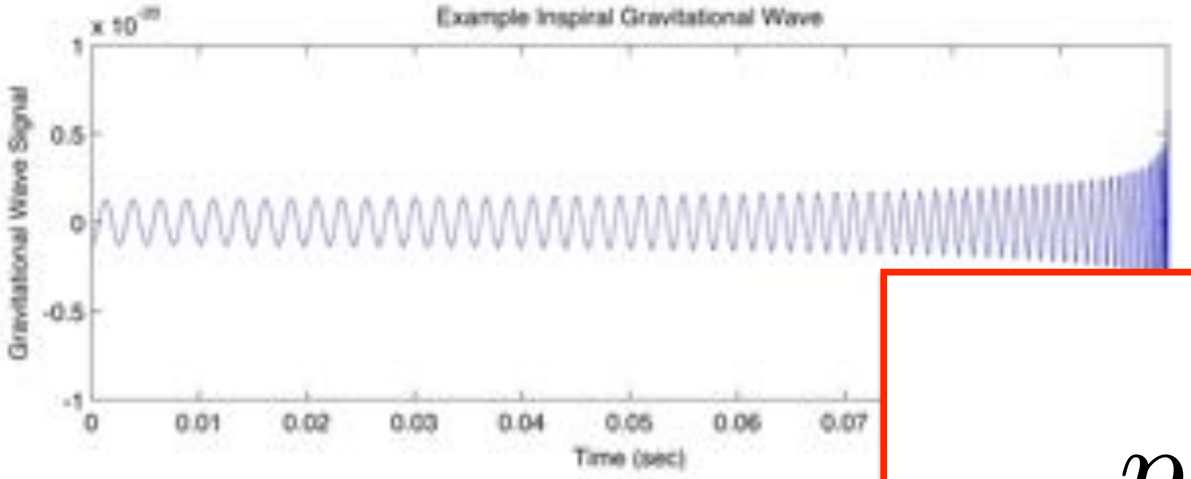
Posterior
probability for
waveform h

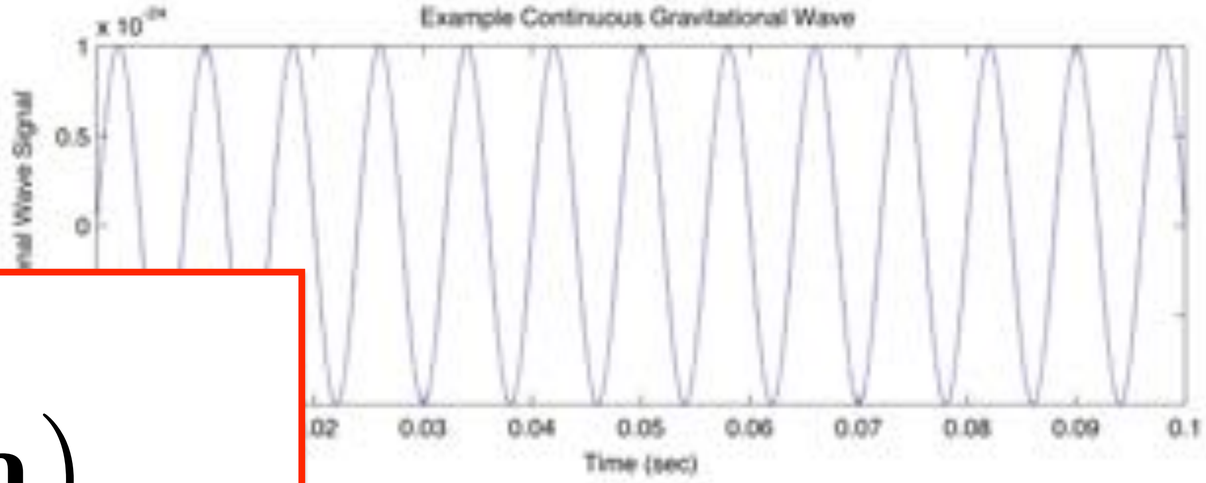
$$p(h|d) = \frac{p(d|h)p(h)}{p(d)}$$

Normalization - model evidence

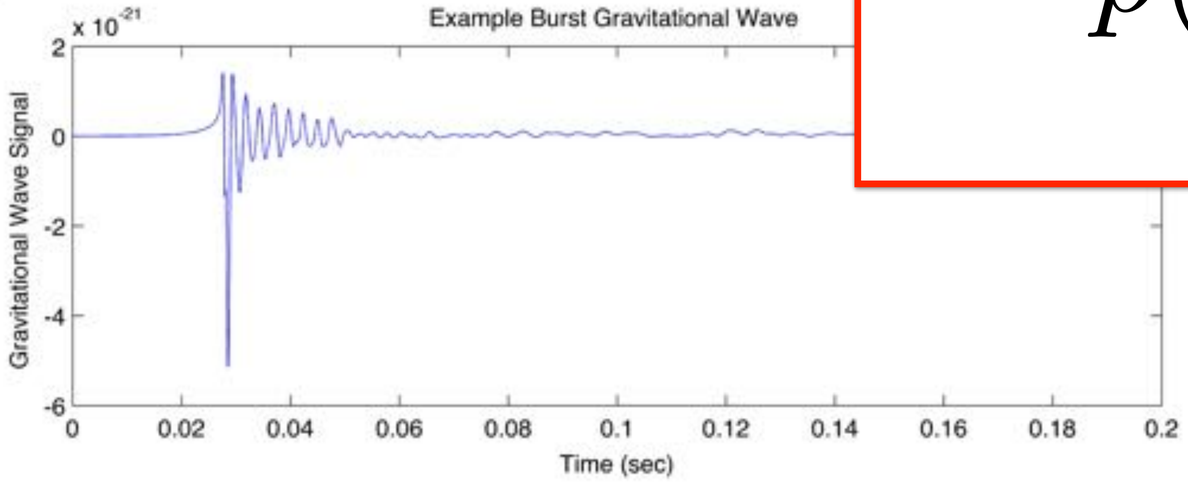
Gravitational wave signal types

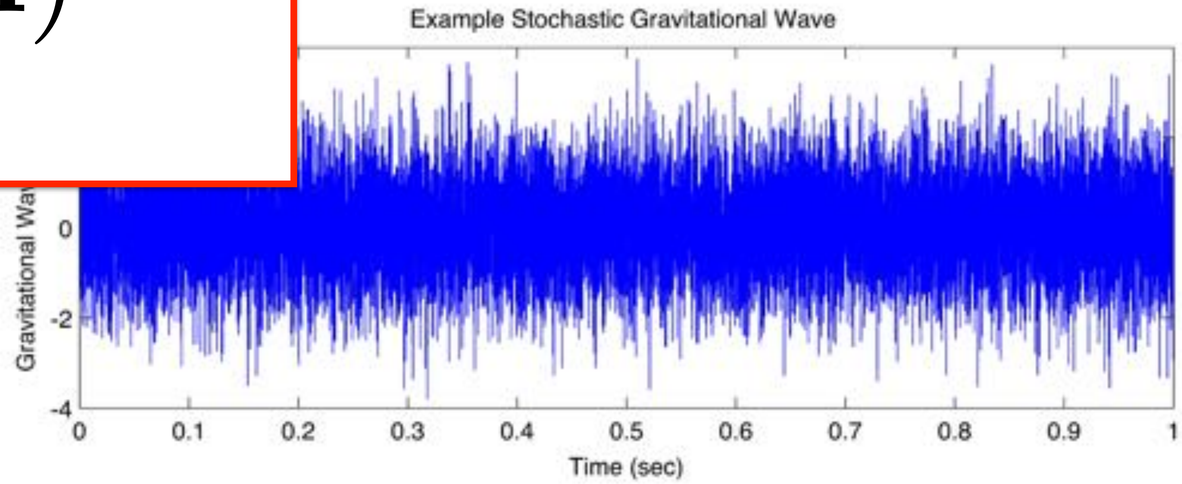
Well modeled - e.g. binary inspiral and merger





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





Poorly modeled - e.g. core collapse supernovae

Stochastic- e.g. phase transition in early universe

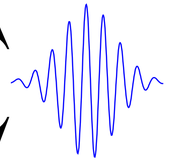
Gravitational wave signal models

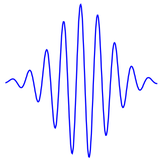
Template based

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

Burst signals

$$p(\mathbf{h}) = \delta(\mathbf{h} - \sum \text{}), \quad p(\text{})$$





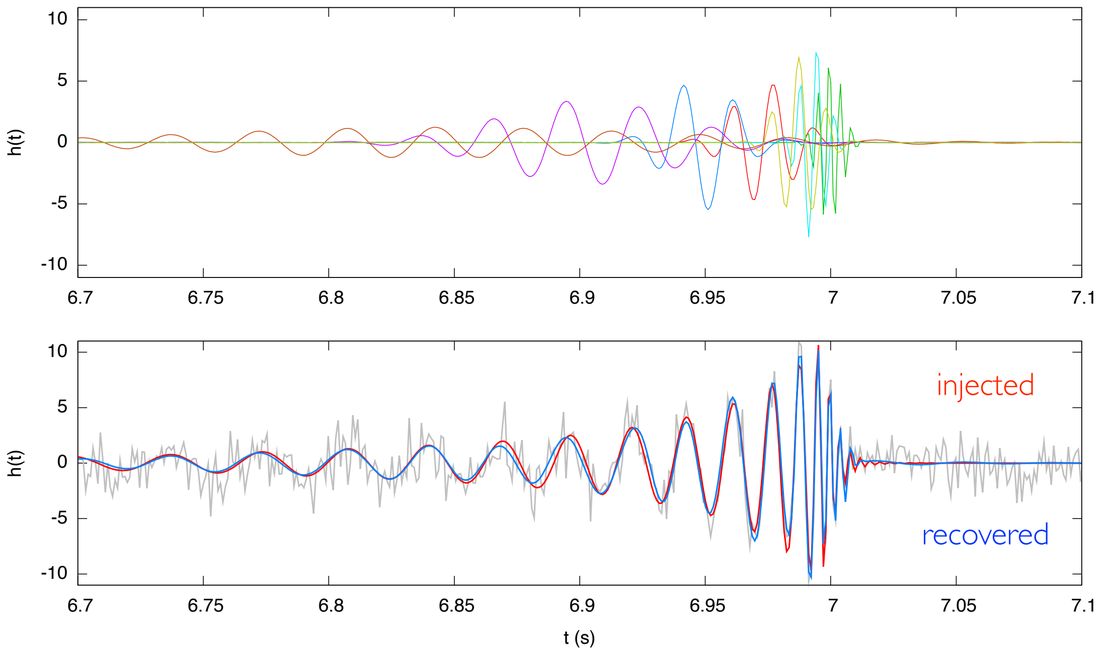
Stochastic signals

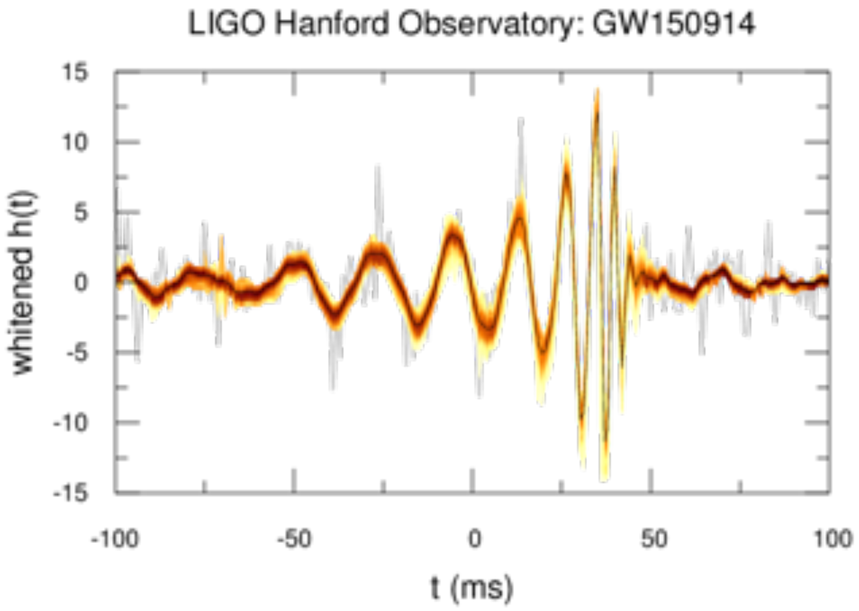
$$p(\mathbf{h}) = \frac{1}{\sqrt{\det(2\pi \mathbf{S_h})}} e^{-\frac{1}{2}(\mathbf{h}^\dagger \mathbf{S_h}^{-1} \mathbf{h})}, \quad p(\mathbf{S_h})$$

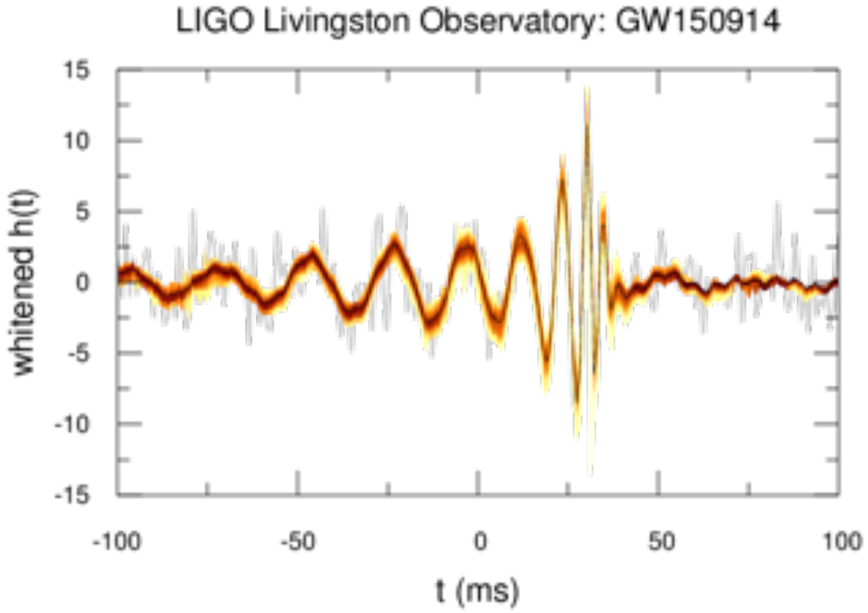
Posterior Distribution for the Waveforms

BayesWave

$$p(\mathbf{h}) = \delta(\mathbf{h} - \sum \text{[waveform]})\, p(\text{[waveform]})$$







Posterior Distribution for the model parameters

Template based models have the strongest priors and hence yield the most sensitive searches

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

Marginal likelihood (hierarchical Bayes)

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

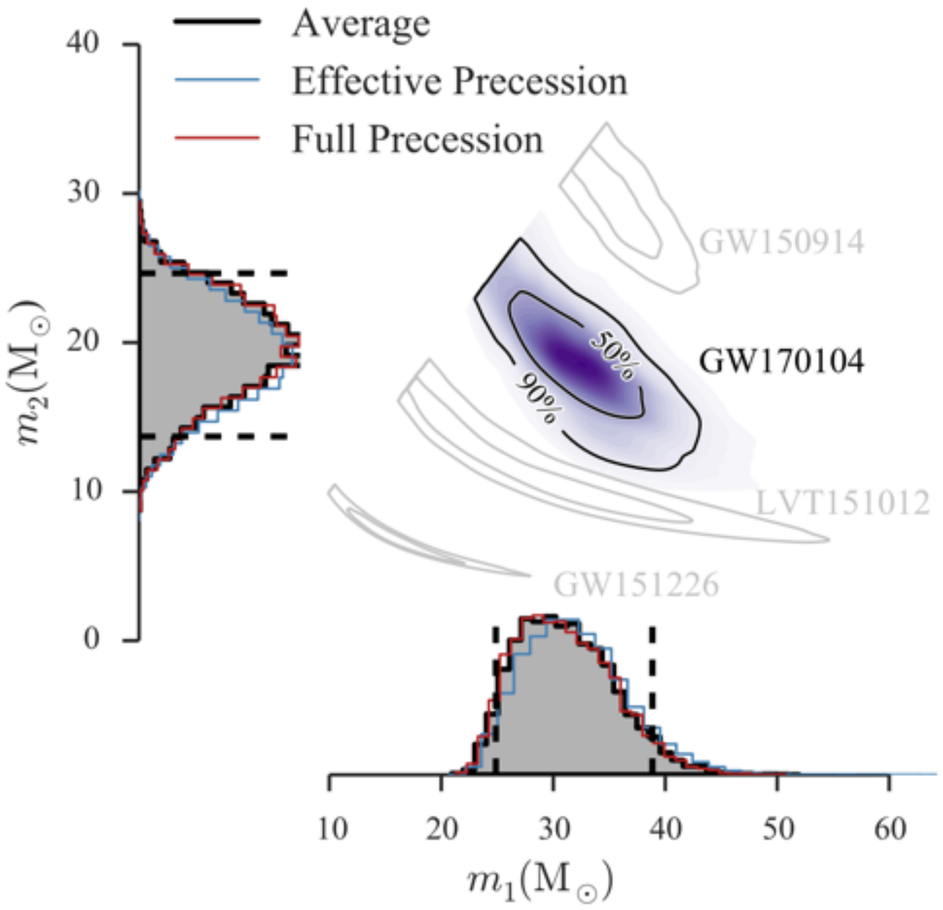
The marginal likelihood and the (hyper) prior on the model parameters then defines the posterior on the model parameters

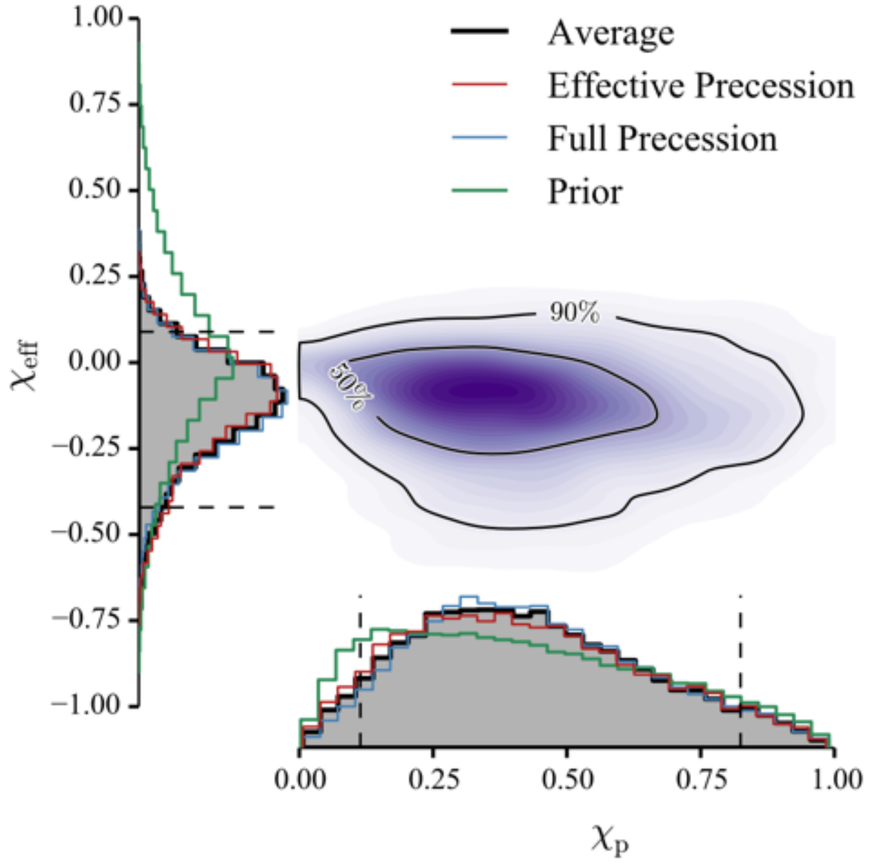
$$p(\vec{\lambda}|\mathbf{d}) = \frac{\overbrace{p(\mathbf{d}|\vec{\lambda})}^{\text{Likelihood}}\overbrace{p(\vec{\lambda})}^{\text{Prior}}}{\underbrace{p(\mathbf{d})}_{\text{Evidence}}}$$

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.

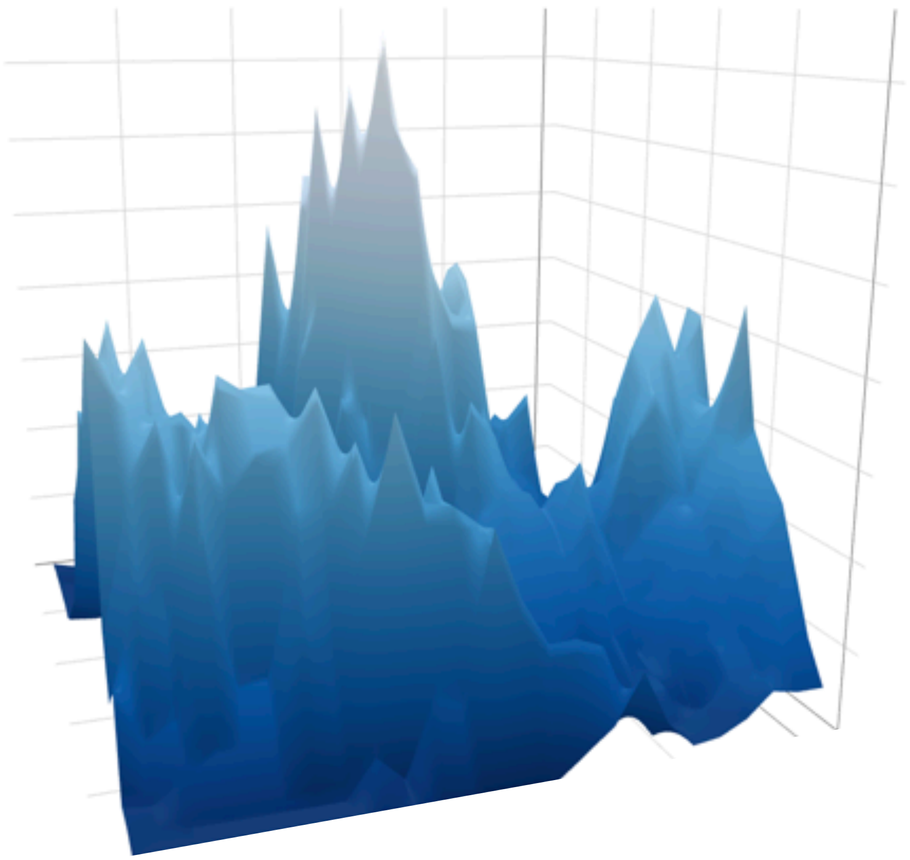
Template based analysis - Parameter Posteriors

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





The Bayesian way - all we need to do is map out the posterior distribution in 17+ dimensions, how hard could that be?



- The likelihood can be multi-modal with narrow peaks
- Waveform generation and/or the likelihood evaluation can be very computationally intensive
- LIGO/Virgo analyses have to sift through vast amounts of data for rare signals
- LISA analyses will have much smaller data sets to contend with, but the data will be contain millions of overlapping sources
- PTA data is unevenly sampled with complicated noise properties

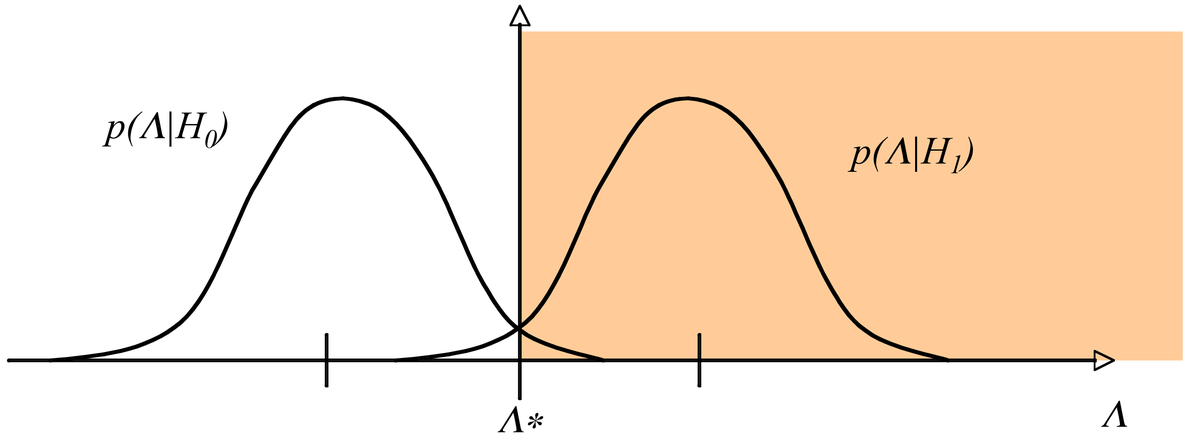
Many analyses start with a search phase followed by Bayesian parameter estimation

Searching for signals (a highly simplified treatment)

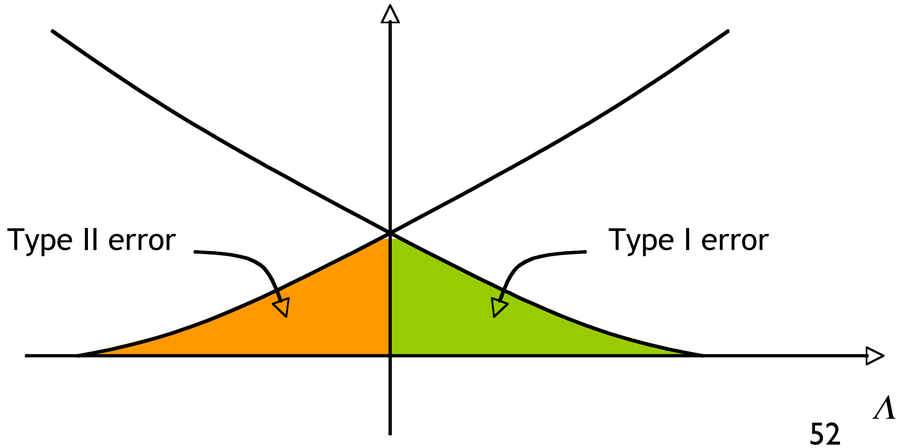
Λ – Detection Statistic

H_0 – Noise Hypothesis

H_1 – Noise + Signal Hypothesis



Set threshold Λ_* such that $\Lambda > \Lambda_*$ favors hypothesis H_1



Type I error - False Alarm

Type II error - False Dismissal

The likelihood ratio statistic

Likelihood for Stationary Gaussian Noise

$$p(\mathbf{d}|\vec{\lambda}) \sim e^{-(\mathbf{d}-\mathbf{h}(\vec{\lambda})|\mathbf{d}-\mathbf{h}(\vec{\lambda}))/2}$$

Noise weighted inner product

$$(\mathbf{a}|\mathbf{b}) = 2\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)}\, df$$

Suppose that we have two hypotheses:

$$H_1 : \text{ A signal with parameters } \vec{\lambda} \text{ is present}$$
$$H_0 : \text{ No signal is present}$$

Likelihood ratio: $\Lambda(\vec{\lambda}) = \dfrac{p(\mathbf{d}|\mathbf{h}(\vec{\lambda}), H_1)}{p(\mathbf{d}, H_0)}$

For Gaussian noise: $\Lambda(\vec{\lambda}) = e^{-(\mathbf{d}|\mathbf{h}) + \frac{1}{2}(\mathbf{h}|\mathbf{h})}$

The ρ - statistic

For a fixed false alarm rate, the false dismissal rate
is minimized by the likelihood ratio statistic (Neyman-Pearson)

$$\Lambda(\vec{\lambda}) = \frac{p(\mathbf{d}|\mathbf{h}(\vec{\lambda}), H_1)}{p(\mathbf{d}, H_0)}$$

The likelihood ratio is maximized over the signal parameters $\vec{\lambda}$

The likelihood ratio can be maximized wrt to the overall amplitude to yield the ρ - statistic:

Writing $h(\vec{\lambda}) = \rho(\vec{\lambda})\,\hat{h}(\vec{\lambda}) \quad \Rightarrow \quad \Lambda(\vec{\lambda}) = e^{\rho(\mathbf{d}|\hat{h}) - \frac{1}{2}\rho^2}$

Maximizing: $\dfrac{\partial \Lambda(\vec{\lambda})}{\partial \rho} = 0$

$$\Rightarrow \quad \rho(\vec{\lambda}) = (\mathbf{d}|\hat{h}(\vec{\lambda}))$$

The ρ-statistic and SNR

The signal-to-noise ratio (SNR) is defined:

$$\begin{aligned}\mathrm{SNR} &= \frac{\text{Expected value when signal present}}{\text{RMS value when signal absent}} \\ &= \frac{E[\rho]}{\sqrt{E[\rho_0^2] - E[\rho_0]^2}} \\ &= (h|\hat{h}) \\ &= \sqrt{(h|h)}\end{aligned}$$

In practice, the detector noise is not perfectly Gaussian, and variants of the ρ - statistic are now used, notably the "new SNR" statistic, introduced by B. Allen Phys.Rev. D71 (2005) 062001

$$\boxed{\mathrm{SNR}^2 = 4\int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)}\,df}$$

Frequentist Detection Threshold

For stationary, Gaussian noise the ρ - statistic is Gaussian distributed.

For the null hypothesis we have $$p_0(\rho) = \frac{1}{\sqrt{2\pi}}\, e^{-\rho^2/2}$$

For the detection hypothesis we have $$p_1(\rho) = \frac{1}{\sqrt{2\pi}}\, e^{-(\rho^2 - \mathrm{SNR}^2)/2}$$

Setting a threshold of ρ_* gives the false alarm and false dismissal probabilities

$$P_{\mathrm{FA}} = \frac{1}{2}\mathrm{erfc}(\rho_*/\sqrt{2})$$
$$P_{\mathrm{FD}} = \frac{1}{2}\mathrm{erfc}((\rho_* - \mathrm{SNR})/\sqrt{2})$$

LIGO/Virgo analyses do not use SNR thresholds, but rather use False Alarm Rate thresholds

$$\mathrm{FAR} = \frac{P_{\mathrm{FA}}}{T_{\mathrm{obs}}}$$

e.g. FAR = One in million years and an observation time of one year

$$P_{\mathrm{FA}} = 10^{-6} \quad \mathrm{aka} \quad 4.9\,\sigma \qquad\qquad \rho_* = 4.8$$

Grid Based Searches

Goal is to lay out a grid in parameter space that is fine enough to catch any signal with some good fraction of the maximum matched filter SNR

The match measures the fractional loss in SNR in recovering a signal with a template and defines a natural metric on parameter space:

$$M(\vec{x},\vec{y}) = \frac{(h(\vec{x})|h(\vec{y}))}{\sqrt{(h(\vec{x})|h(\vec{x}))(h(\vec{y})|h(\vec{y}))}}$$

Taylor expanding $\quad M(\vec{x},\vec{x}+\Delta\vec{x}) = 1 - g_{ij}\Delta x^i \Delta x^j + \ldots$

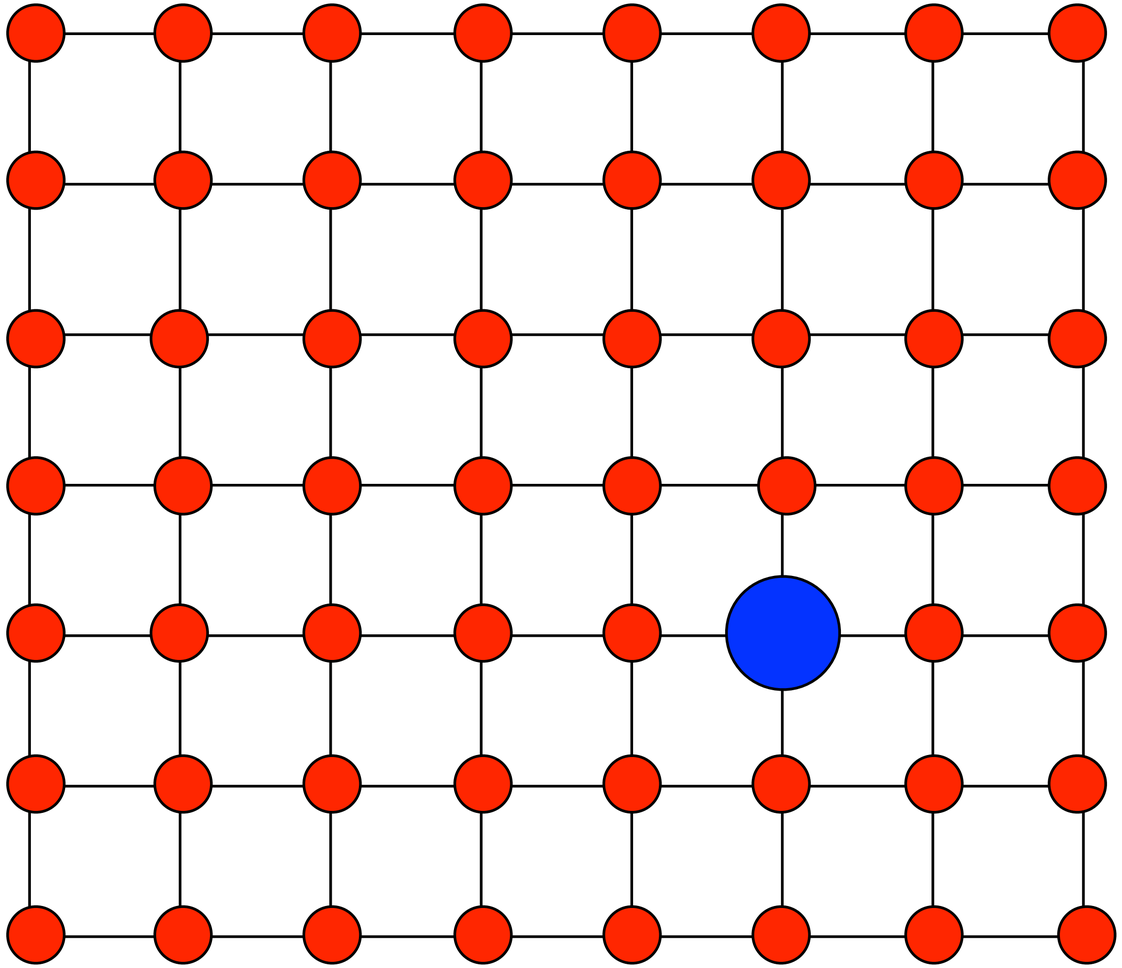
where $\quad g_{ij} = \frac{(h_{,i}|h_{,j})}{(h|h)} - \frac{(h|h_{,i})(h|h_{,j})}{(h|h)^2}$ (Owen Metric)

Number of templates (for a hypercube lattice in D dimensions)

$$N = \frac{V}{\Delta V} = \frac{\int d^D x \sqrt{g}}{(2\sqrt{(1-M_{\rm min})}/D)^D}$$

Cost grows geometrically with D for any lattice

LIGO Style Grid Searches



Typically 2-3 dimensional, 10,000's points

Reducing the cost of a search

In most cases it is possible to analytically maximize over 3 or more parameters

Distance:

The unit normalized template $\hat{h}$ defines a reference distance $\bar{D}$

Scaling this template to distance D gives

$$h = \frac{\bar{D}}{D}\,\hat{h}$$

The distance is then estimated from the data as

$$D = \frac{\bar{D}}{(d|\hat{h})}$$

Reducing the cost of a search

Phase Offset:

Generate two templates $h(\phi = 0)$ and $h(\phi = \pi/2)$

Then $$(d|h)_{\max\,\phi} = \sqrt{(d|h(0))^2 + (d|h(\pi/2))^2}$$

Easy to see this in the Fourier domain.

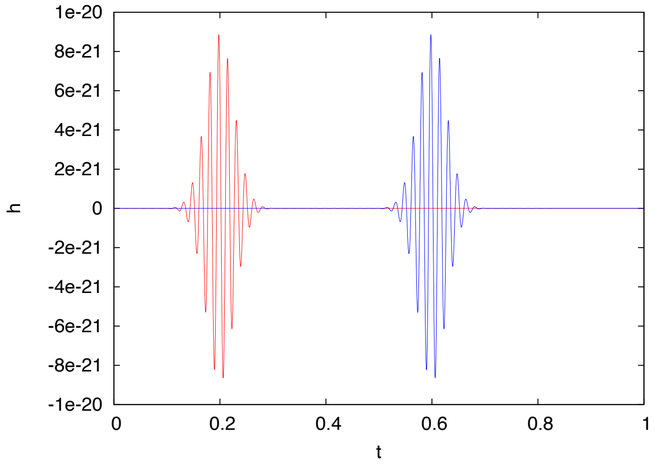
Suppose $\tilde{d} = \tilde{h}_0\, e^{i\phi}$, then

$$(d|h(0)) = (h_0|h_0)\cos\phi$$

$$(d|h(\pi/2)) = (h_0|h_0)\sin\phi$$

Reducing the cost of a search

Time Offset:



Fourier transform treats time as periodic - use this to our advantage

Compute the inverse Fourier transform of the product of the Fourier transforms:

$$(d|h)(\Delta t) = 4 \int \frac{\tilde{d}^*(f)\tilde{h}(f)}{S(f)} e^{2\pi i f \Delta t} \, df$$

Then if the template and data differ by a time shift: $d(t) = h(t - t_0)$

$$(d|h)_{\max t} = (d|h)(\Delta t = t_0)$$

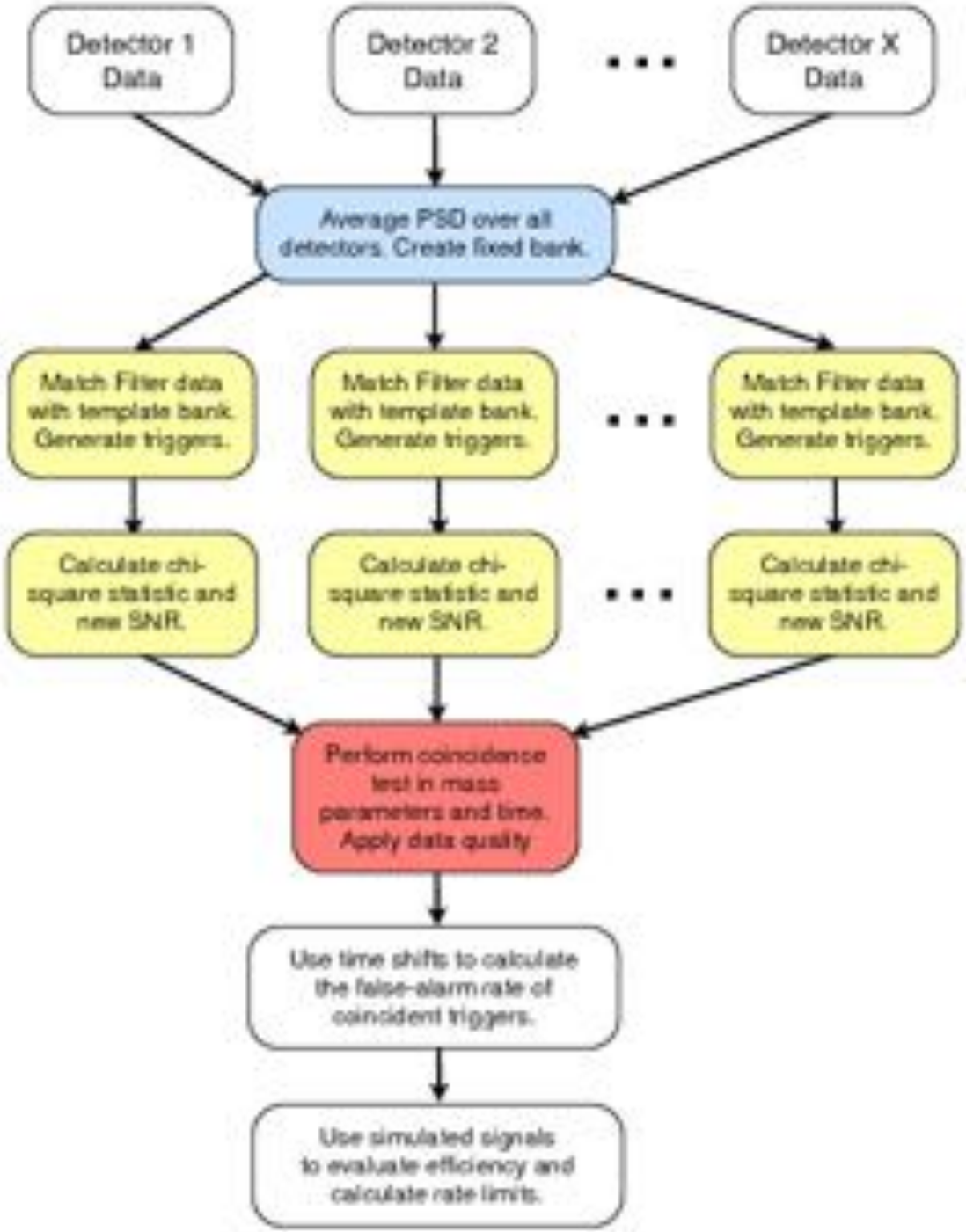
Reducing the cost of a search - putting it all together

$$z(t, \overrightarrow{\eta}) = 4 \int \frac{d(f)\hat{h}^*(f, \overrightarrow{\eta})}{S(f)} e^{2\pi i f t}$$

Time $$\rho(t_0, \overrightarrow{\eta}) = \max_t [\,|z(t, \overrightarrow{\eta})|\,]$$

Phase $$\varphi(\overrightarrow{\eta}) = \arg\{z(t_0, \overrightarrow{\eta})\}$$

Distance $$D(\overrightarrow{\eta}) = \bar{D}\,\rho(t_0, \overrightarrow{\eta})$$



Workflow for pyCBC search

Template bank constructed

Matched filtering is done per-detector (not coherent)

Detection statistic computed (“new SNR”)

Coincidence in time/mass enforced
Data quality vetoes applied

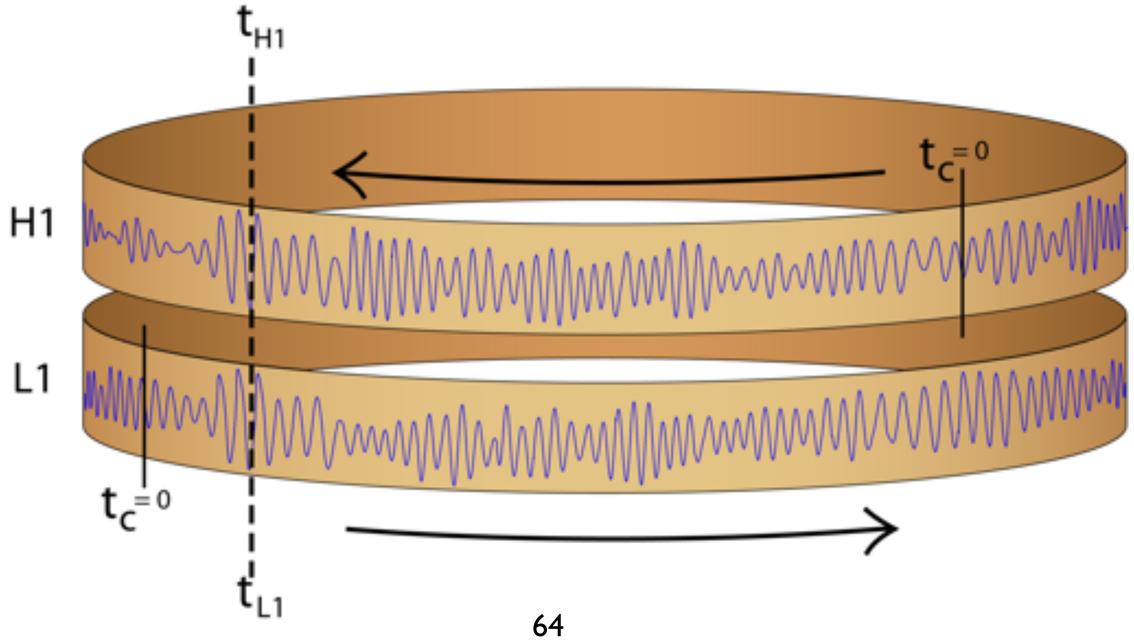
Monte Carlo background to compute FAR vs new SNR

Contending with non-stationary, non-Gaussian noise

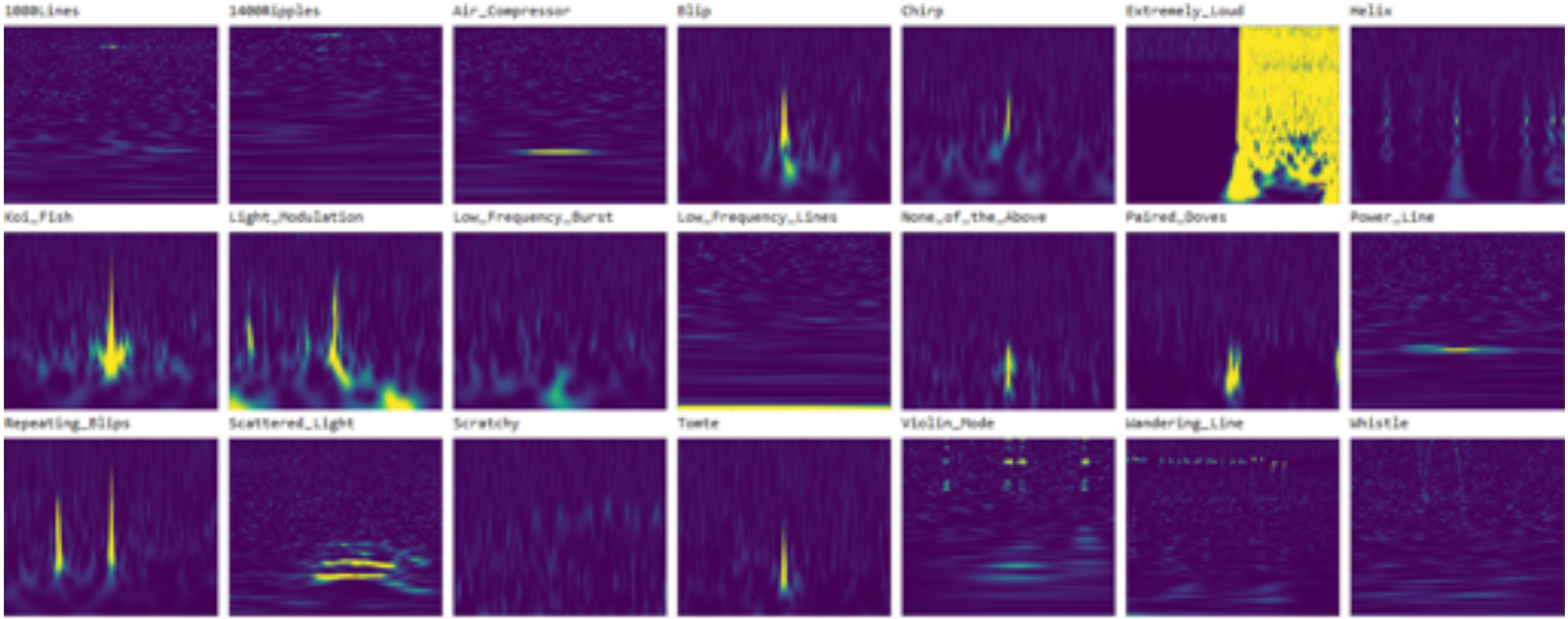
Non-stationary: → Adiabatic drifts in the PSD — - work with short data segments

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

Non-Gaussian: → Glitches

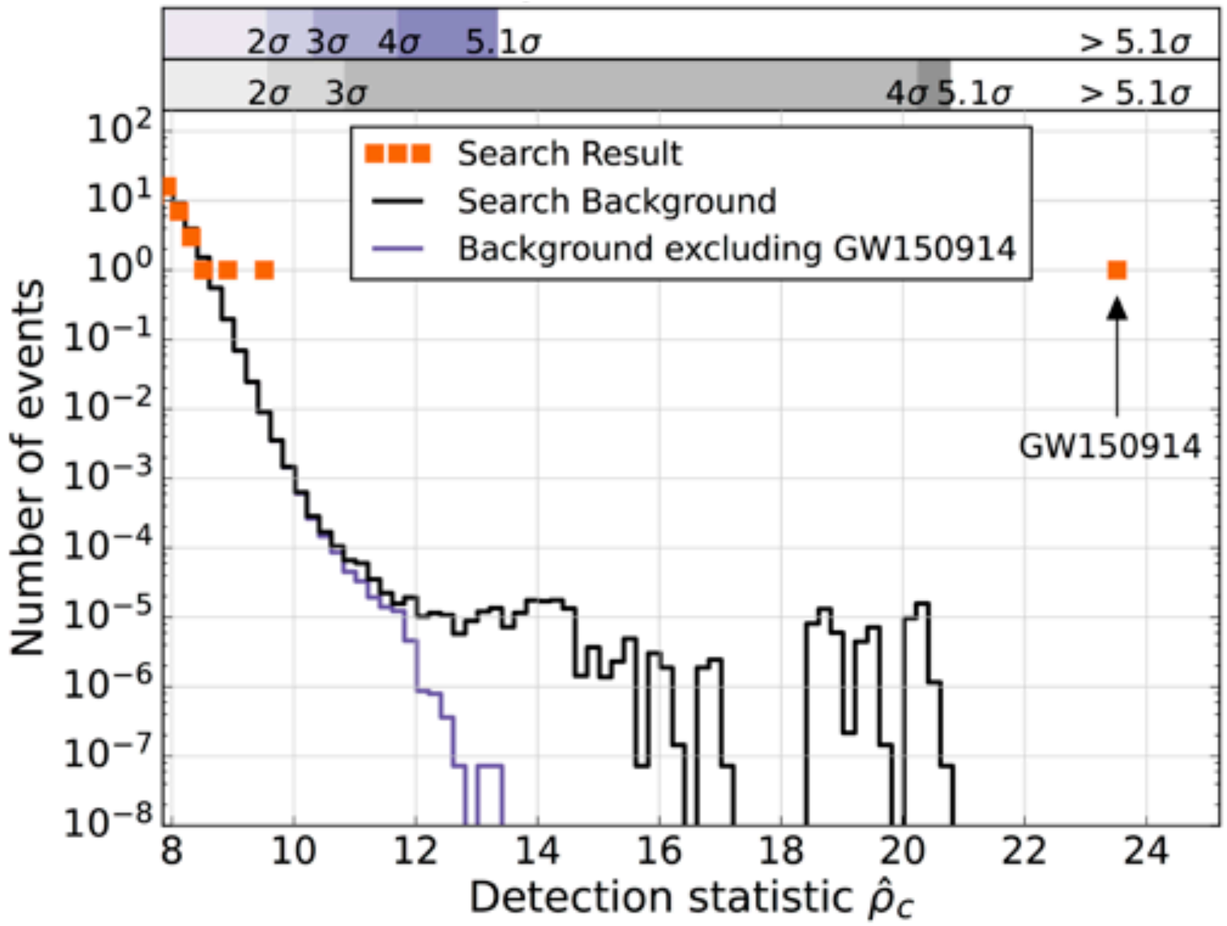


Non-Gaussian Noise Transients



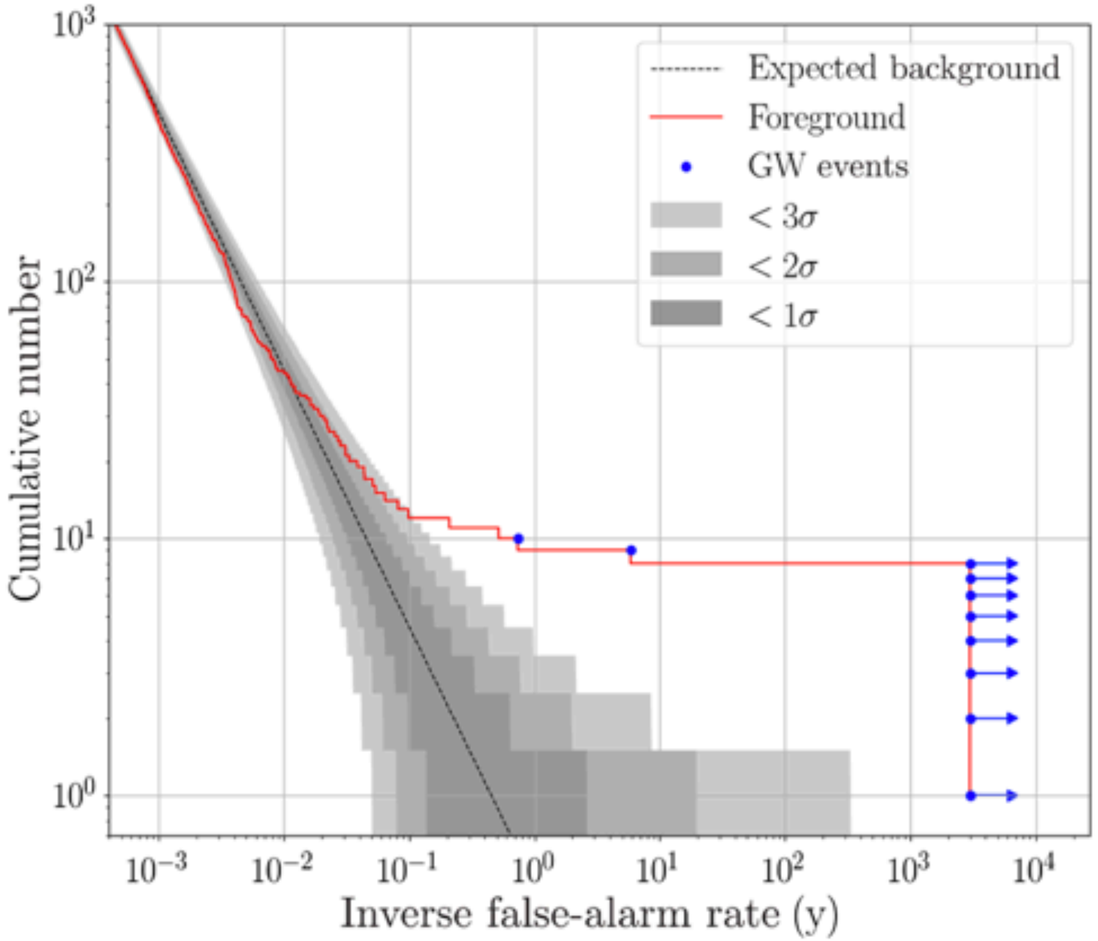
Search results

First detection



"new SNR"

GWTC-1 (O1 & O2)



Gravitational wave data analysis

Neil J. Cornish

extreme Gravity Institute

MONTANA STATE UNIVERSITY

Bayesian Inference

- Bayesian Probability Theory
- Bayesian Learning
- Model Selection
- Markov Chain Monte Carlo
- Trans-dimensional Inference

Bayesian Parameter Estimation

Degree of belief interpretation of probability - the natural expression of the scientific method

Initial Understanding	$\Rightarrow$	New Observations	$\Rightarrow$	Updated Understanding
$p(\vec{x})$		$p(d\|\vec{x})$		$p(\vec{x}\|d)$
Prior	$\Rightarrow$	Likelihood	$\Rightarrow$	Posterior

Bayes' Theorem

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

Normalization factor is the marginal likelihood or ***evidence***

$$p(d) = \int p(\vec{x})p(d|\vec{x})\, d\vec{x}$$

Bayesian Probability Theory

The posterior distribution fully characterizes the model.

E.g. expectation values

$$E[x^i] = \int x^i \, p(\vec{x}|d) \, d\vec{x}$$

E.g. single parameter probability distributions

$$p(x^i|d) = \int p(\vec{x}|d) \, dx^1 dx^2 \ldots dx^{i-1} dx^{i+1} \ldots dx^D$$

E.g. quantile regions, such as 90%

$$\begin{aligned} 0.05 &= \int^{x_1} p(x|d) \, dx \\ 0.9 &= \int_{x_1}^{x_2} p(x|d) \, dx \end{aligned}$$

Bayesian Learning

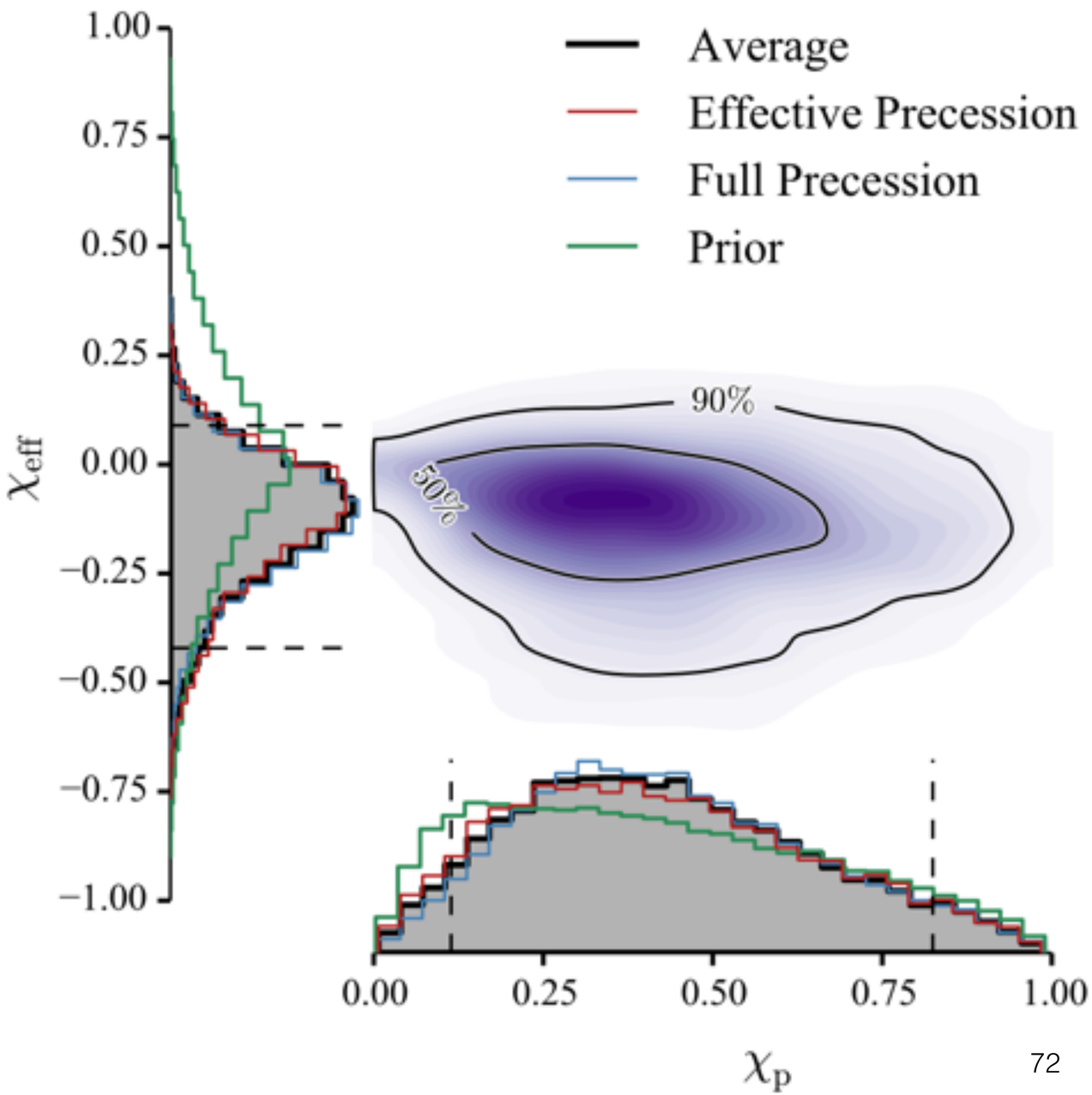
"The (Bayesian) theory of probabilities is basically just common sense reduced to calculus" - Laplace

"Today's posterior is tomorrow's prior" - Lindley

The amount we learn from the data can be measured in bits, and can be computed in terms of the Kullback–Leibler divergence

$$D_{KL} = \int p(\vec{x}|d) \, \log_2 \left(\frac{p(\vec{x}|d)}{p(\vec{x})} \right) \, d\vec{x} \quad [\text{bits}]$$

Bayesian Learning



In this example the prior and posterior distribution for the spins are not so different, especially for the precession. Means that we didn't learn much about the spin.

$$D_{KL} = \int p(\vec{x}|d)\, \log_2 \left(\frac{p(\vec{x}|d)}{p(\vec{x})} \right) d\vec{x} \quad [\text{bits}]$$

Bayesian Model Selection

Probability of Model M: $$p(M|d) \propto \underbrace{p(M)}_{\text{Prior Probability of M}} \underbrace{p(d|M)}_{\text{Evidence for M}}$$

Odds Ratio:

$$\begin{aligned} O_{ij} &= \frac{p(M_i|d)}{p(M_j|d)} \\ &= \frac{p(M_i)}{p(M_j)} \frac{p(d|M_i)}{p(d|M_j)} \\ &= \text{Prior Odds Ratio} \times \text{Bayes Factor} \end{aligned}$$

More on how we compute the Bayes Factor later...

Bayesian Machinery: Markov Chain Monte Carlo

Bayes' Theorem $$p(\vec{x}|d) = \frac{p(\vec{x})p(d|\vec{x})}{p(d)}$$

Marginal likelihood or evidence $$p(d) = \int p(\vec{x})p(d|\vec{x})\, d\vec{x}$$

We know how to compute the prior and the likelihood. The difficulty lies in computing the evidence.

The MCMC technique, introduced by Metropolis and developed by Hastings, allows us to simulated samples from the posterior distribution directly, without having to compute the evidence.

It is possible to compute the evidence using augmented MCMC techniques. Another powerful technique for computing the evidence and the posterior distributions is *Nested Sampling*

Bayesian Inference

Prior

$p(h|M)$

Likelihood

$p(d|h)$

$p(d|M)$

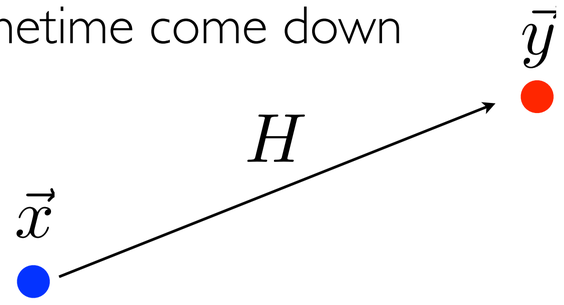
Evidence

$p(h|d, M)$

Posterior

Markov Chain Monte Carlo

Always go up,
Sometime come down



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

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

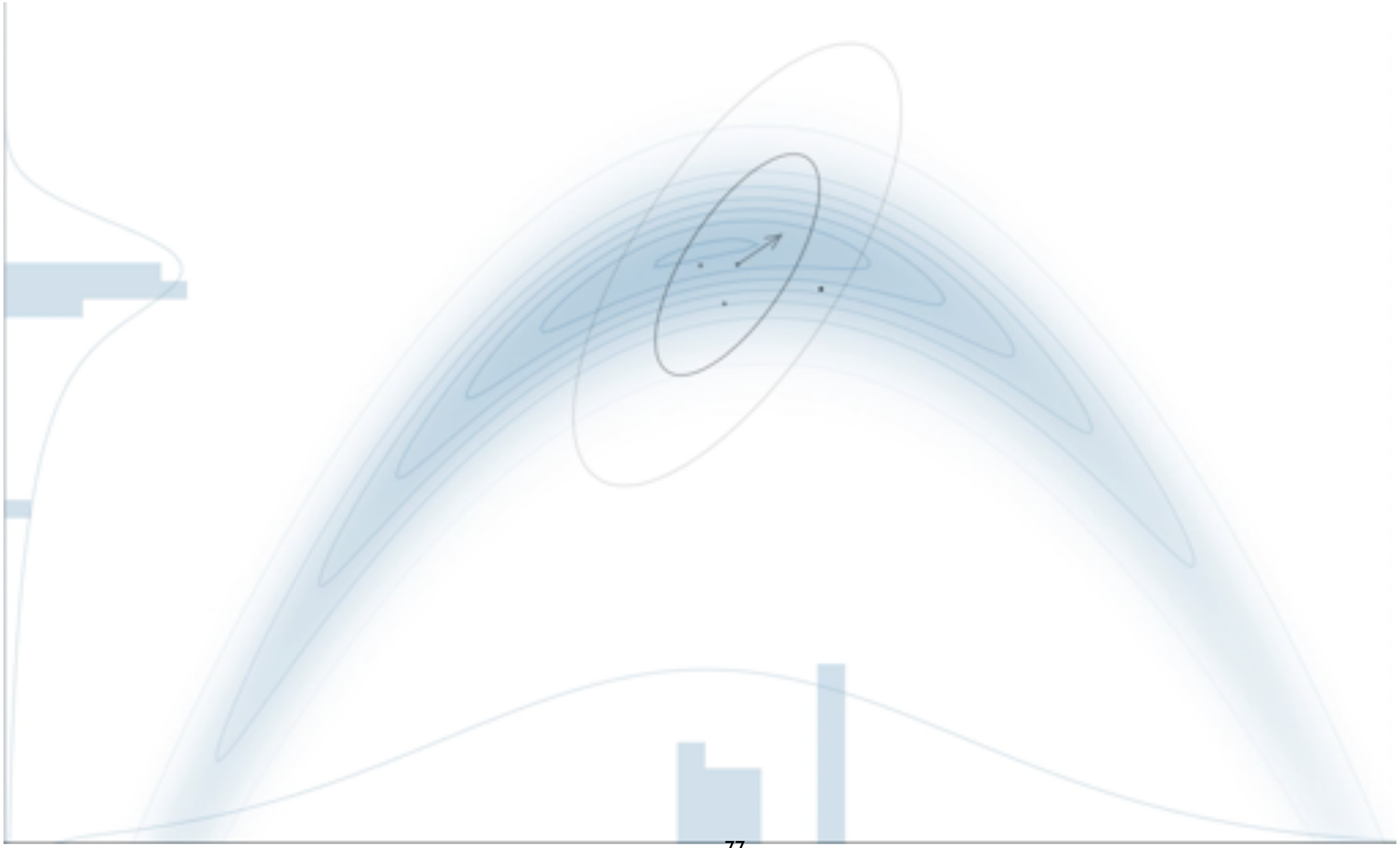
Prior

Likelihood

Proposal

Transition Probability
(Metropolis-Hastings)

https://chi-feng.github.io/mcmc-demo/app.html?algorithm=RandomWalkMH&target=banana



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

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

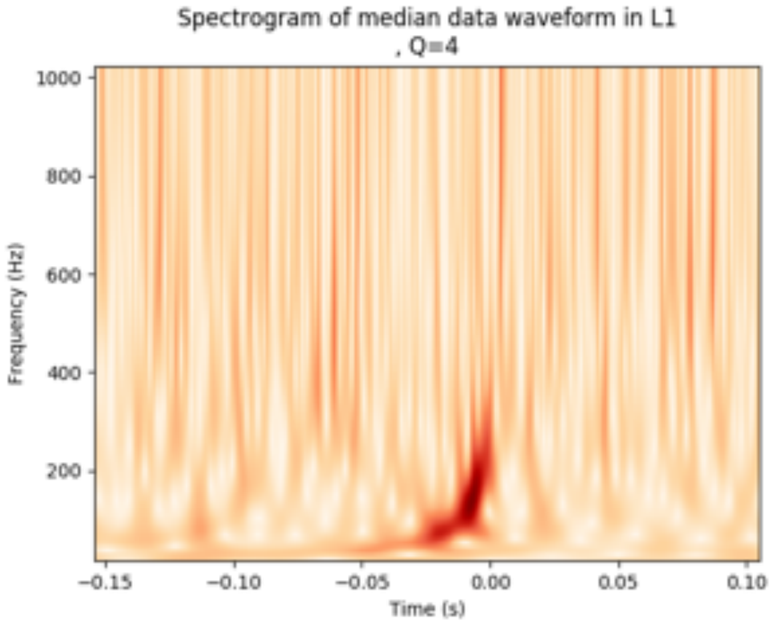
Propose jumps along eigendirections of **K**, scaled by eigenvalues

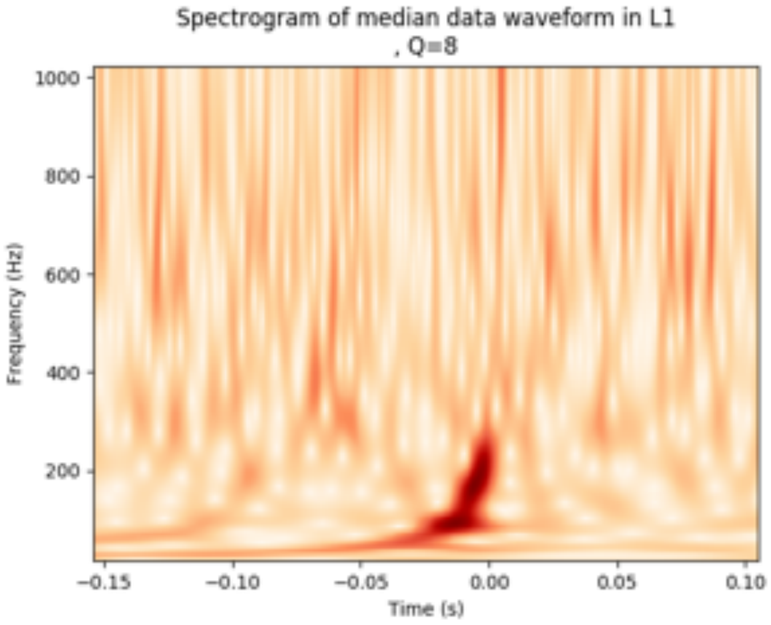
Global likelihood maps

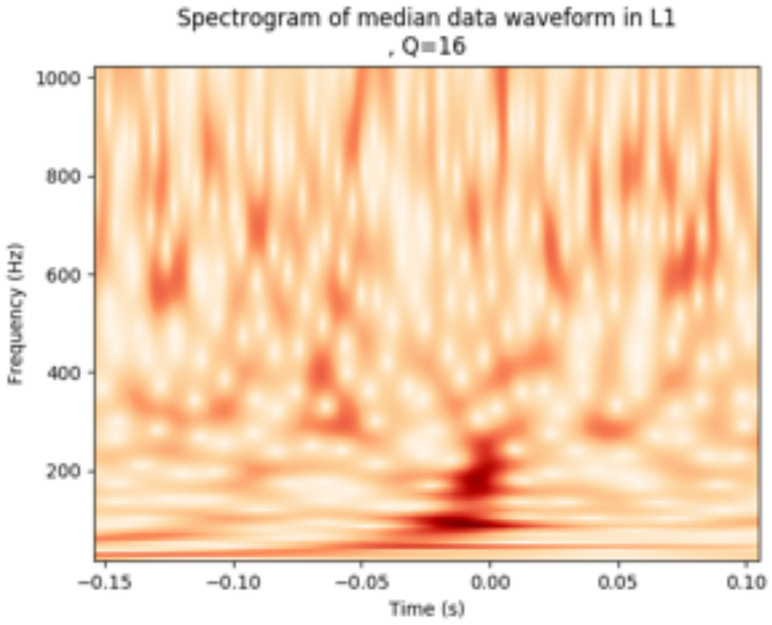
Use a Non-Markovian Pilot search (hill climbers, simulated annealing, genetic algorithms etc) to crudely map the posterior/ likelihood and use this as a proposal distribution for a Markovian follow-up [Littenberg & Cornish, PRD 80, 063007, (2009)]

Time-frequency maps, Maximized likelihood maps

BayesWave Global Map Proposal

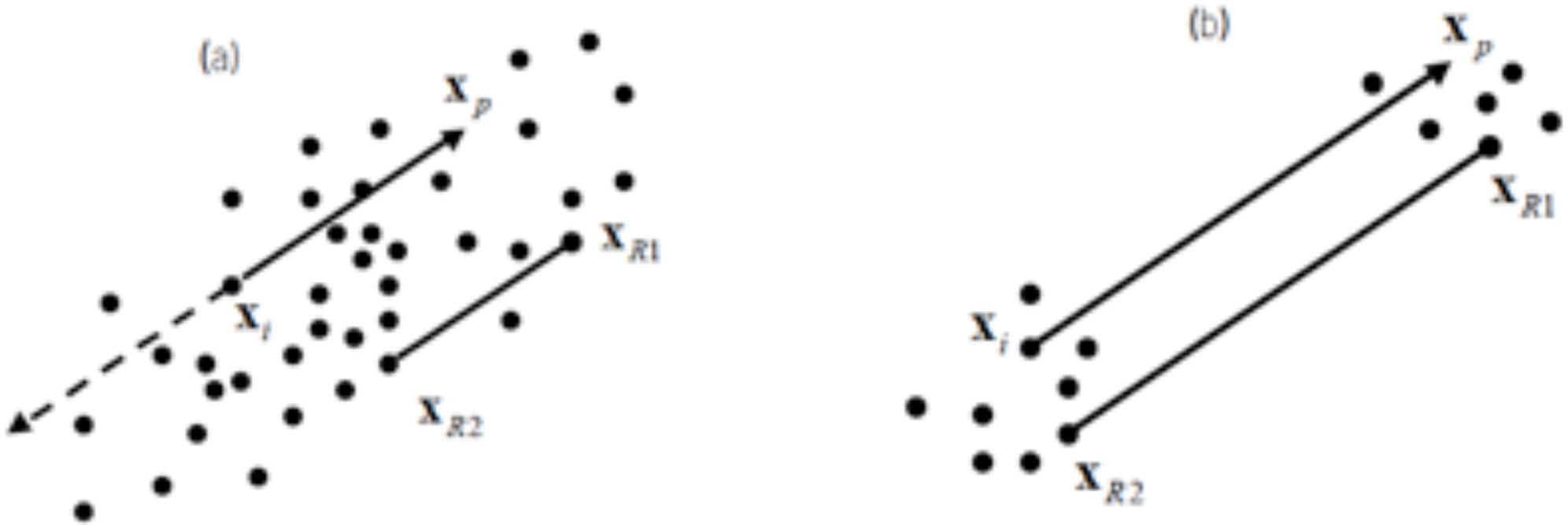






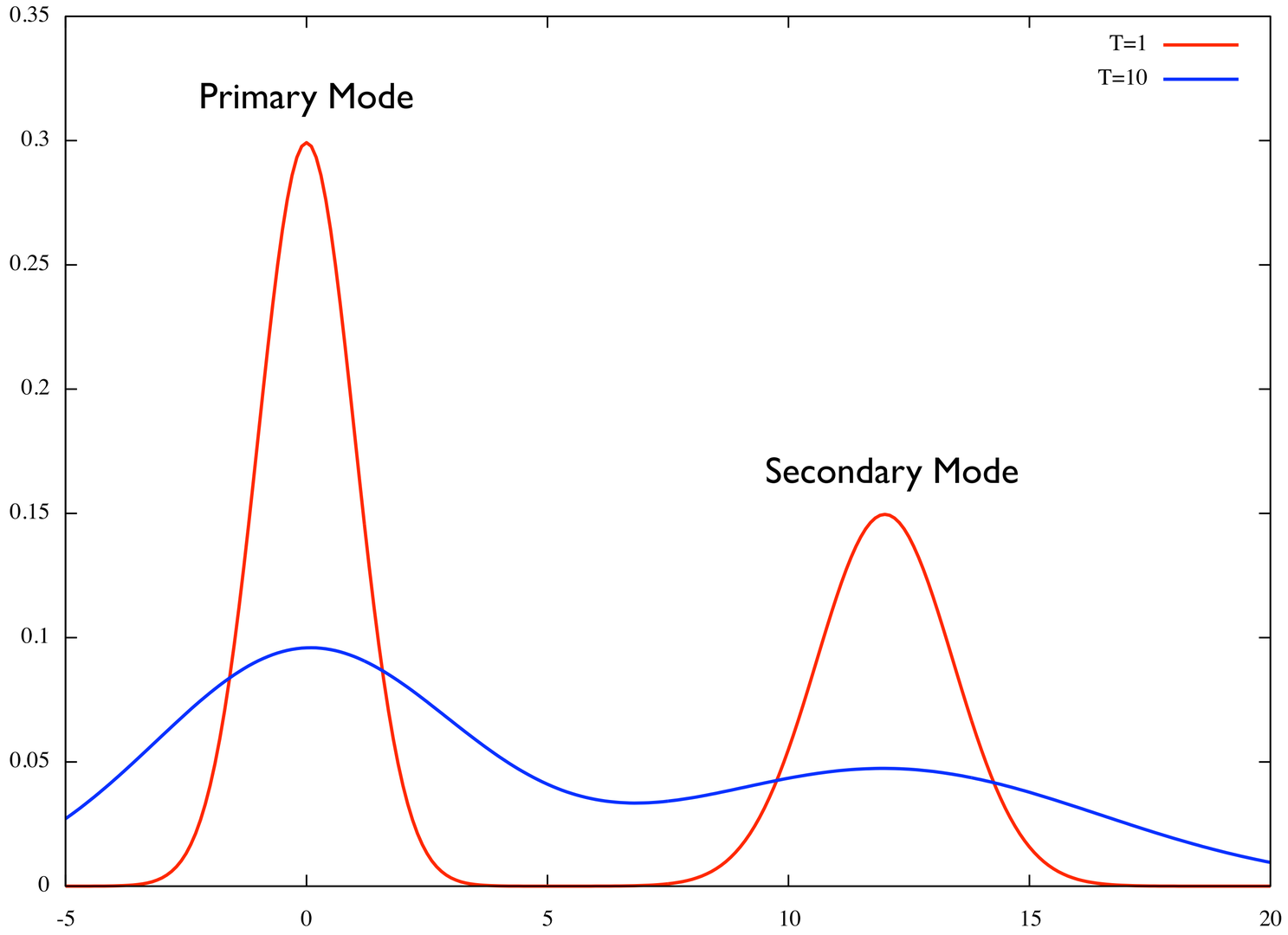
Proposal Distributions

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

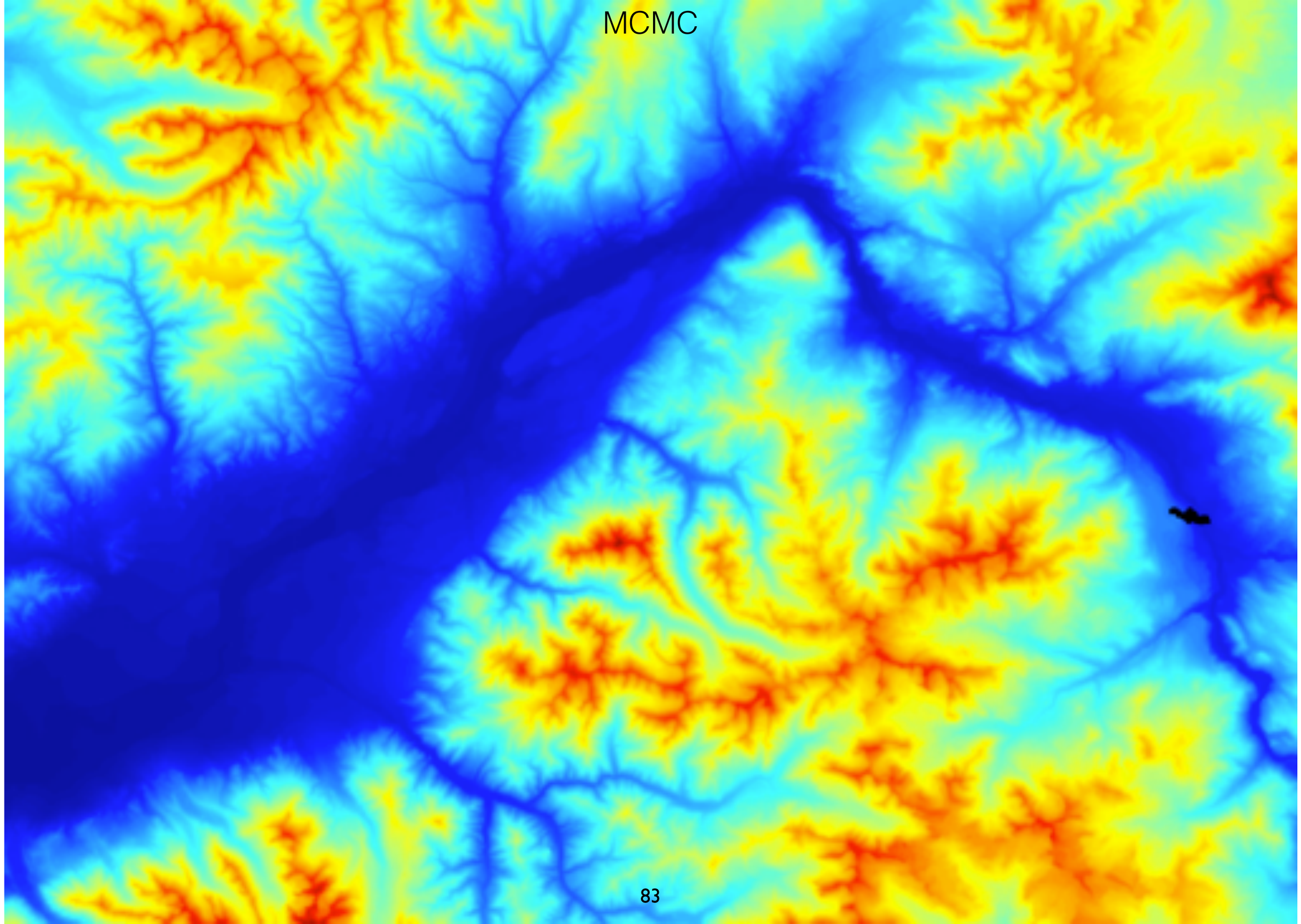
$$\pi(\vec{\lambda}|\mathbf{d})_T = p(\mathbf{d}|\vec{\lambda})^{1/T} p(\vec{\lambda})$$

Compute model evidence

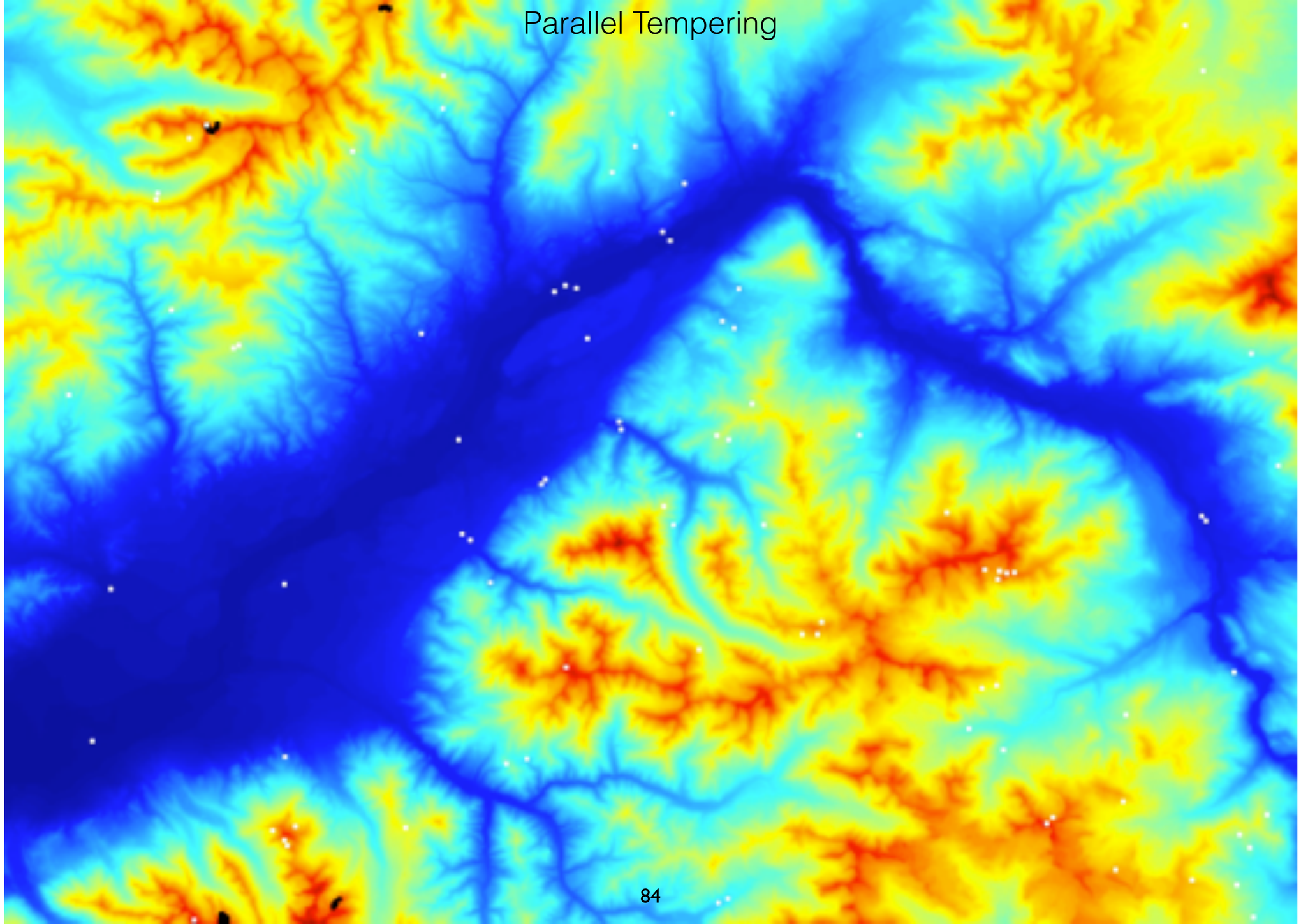
$$\log p(\mathbf{d}) = \int_0^1 \mathbb{E}[\log p(\mathbf{d}|\vec{\lambda})]_\beta \, d\beta$$

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

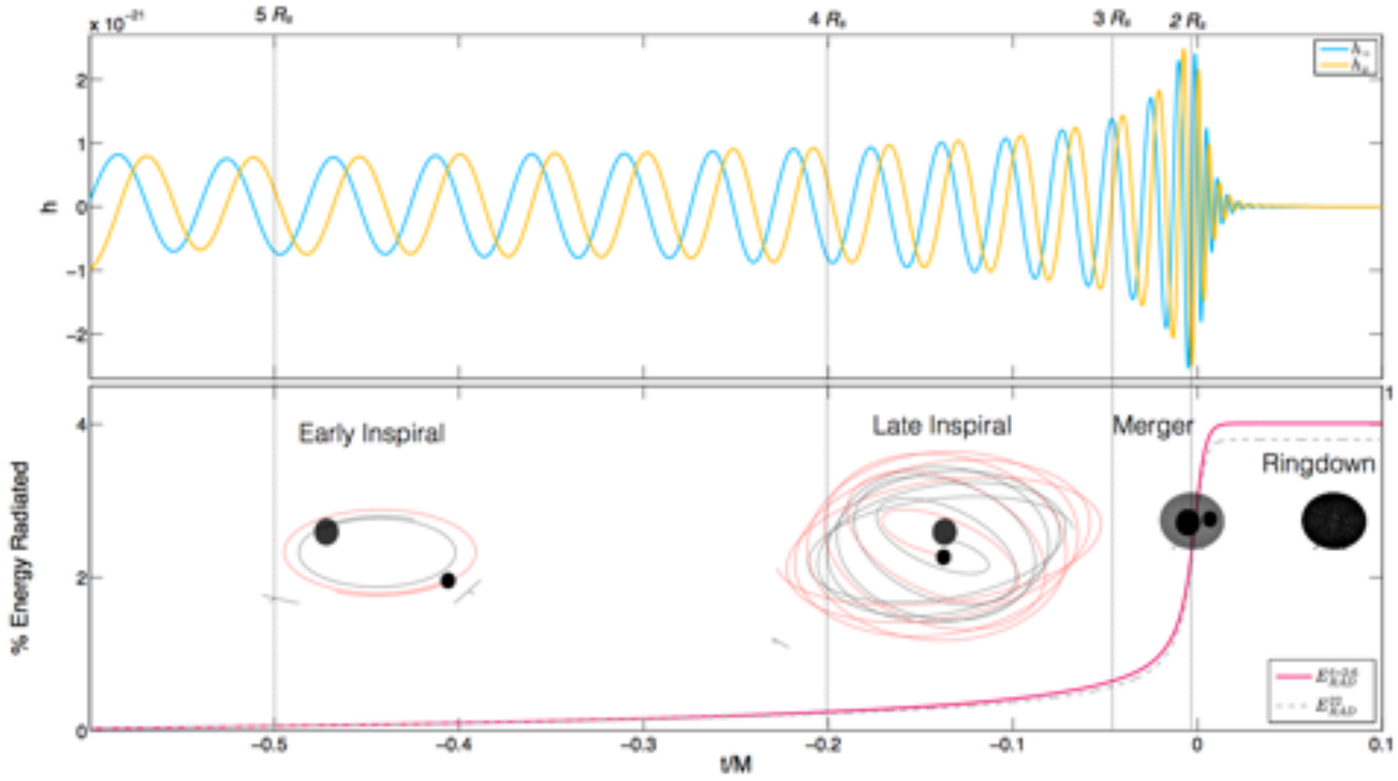
MCMC



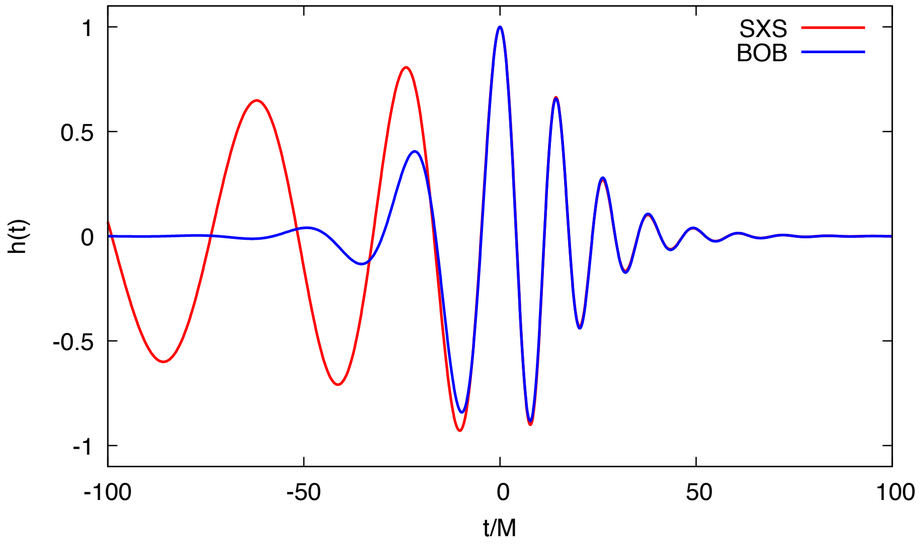
Parallel Tempering



How information is encoded in GW signals



[S. McWilliams PRL 122, 191102 (2019)]



Merger-ringdown encodes final mass and spin

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

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

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

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

$$A_2(t) = A_* \,\mathrm{sech}(t/\tau)$$

$$f_2(t) = \left(\frac{(f_\infty^4 + f_0^4) - (f_\infty^4 - f_0^4)\tanh(t/\tau)}{2} \right)^{1/4}$$

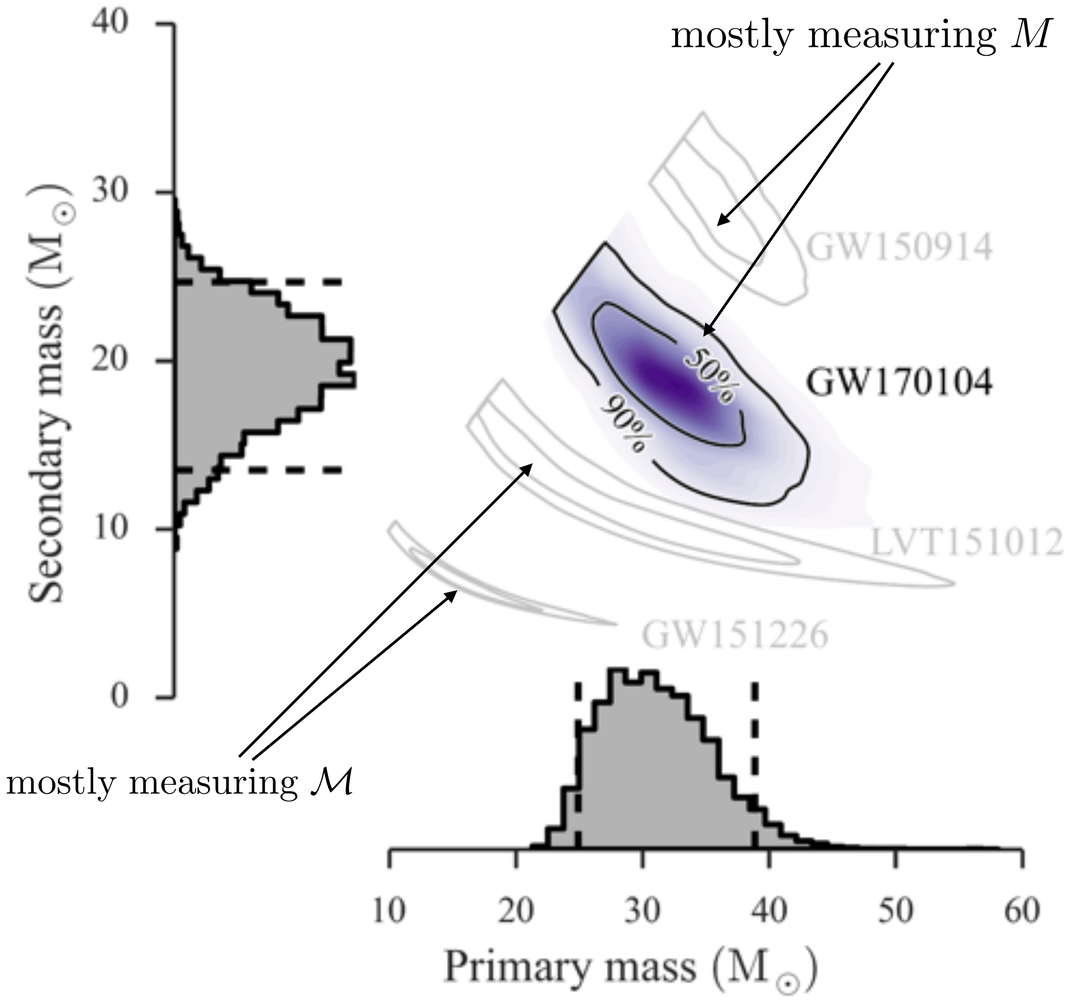
$$\tau = j(a_f)\, M_f \qquad f_\infty = \frac{g(a_f)}{M_f}$$

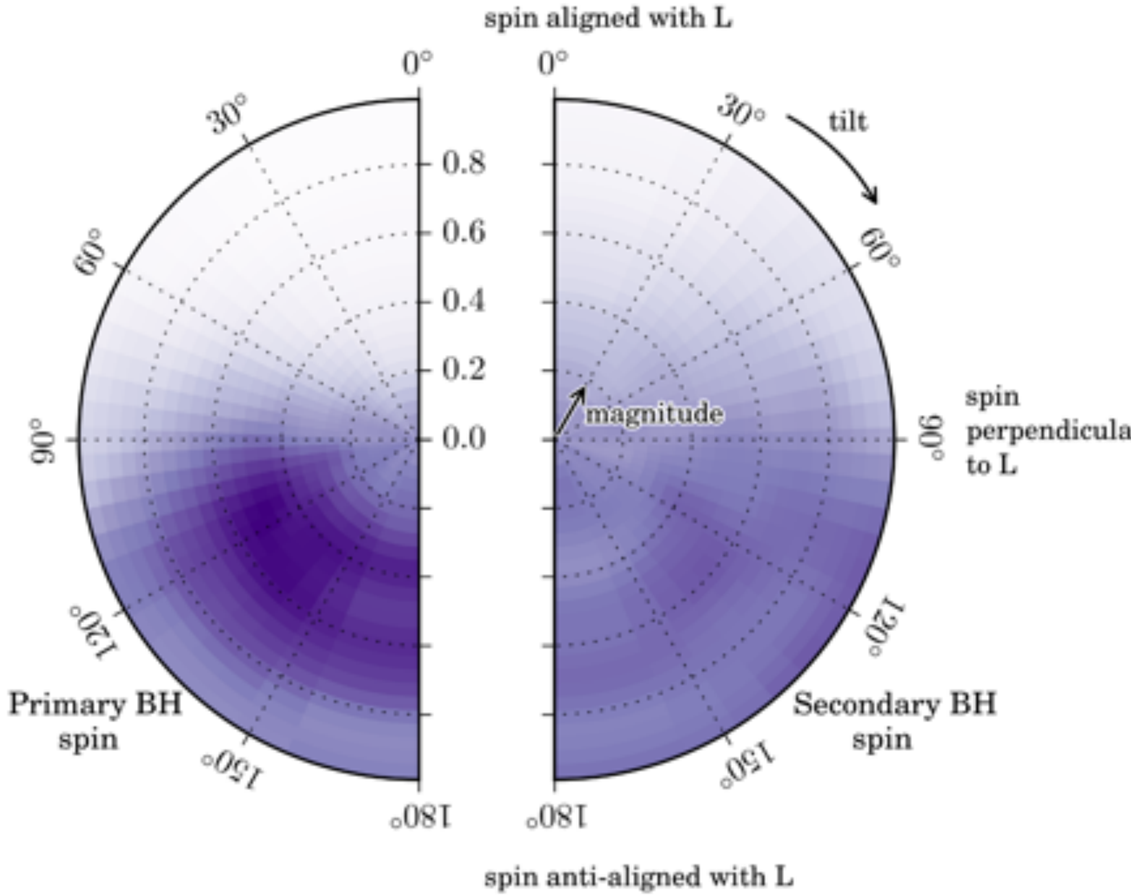
Post-Newtonian Expansion

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

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

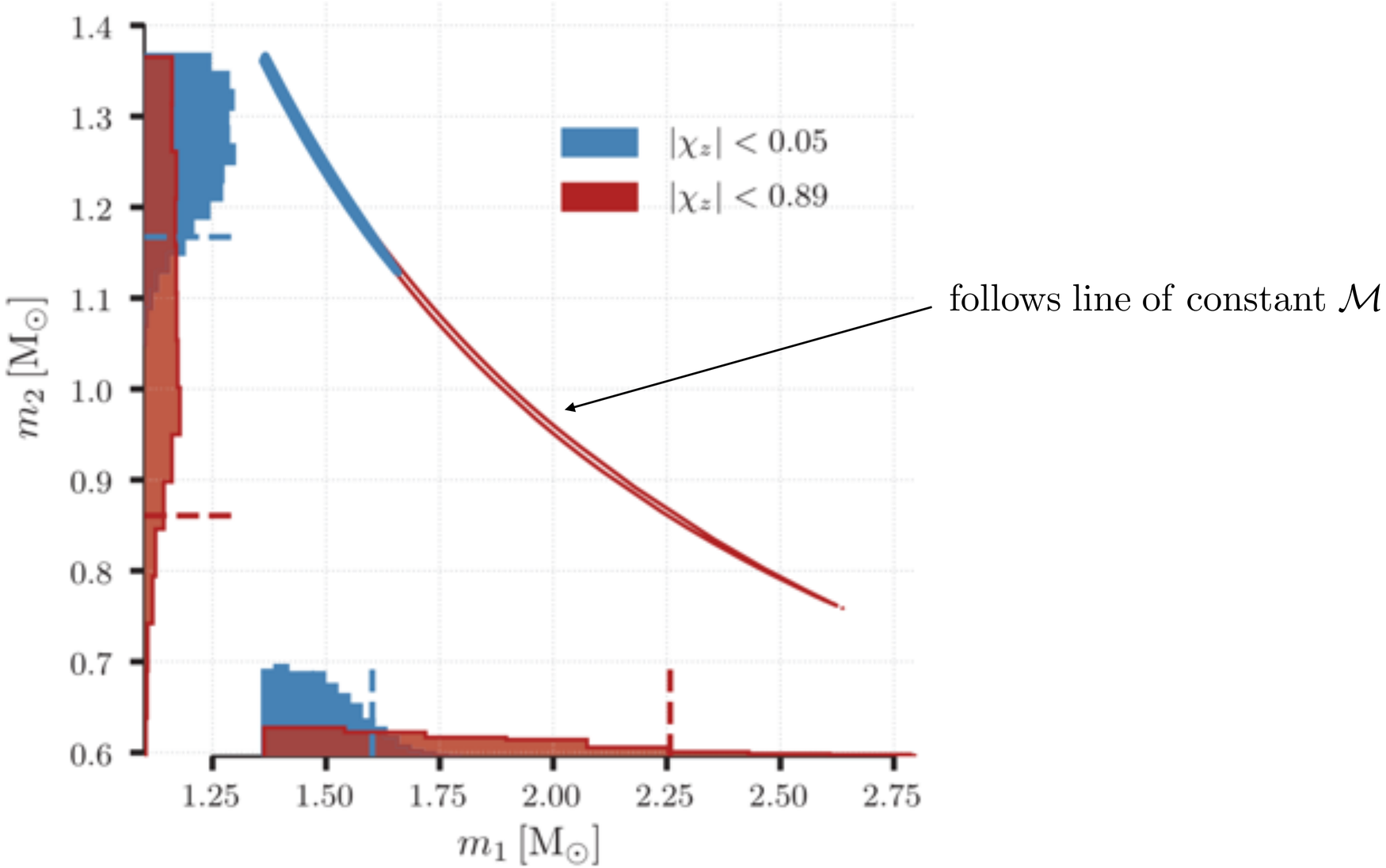
Parameter estimation





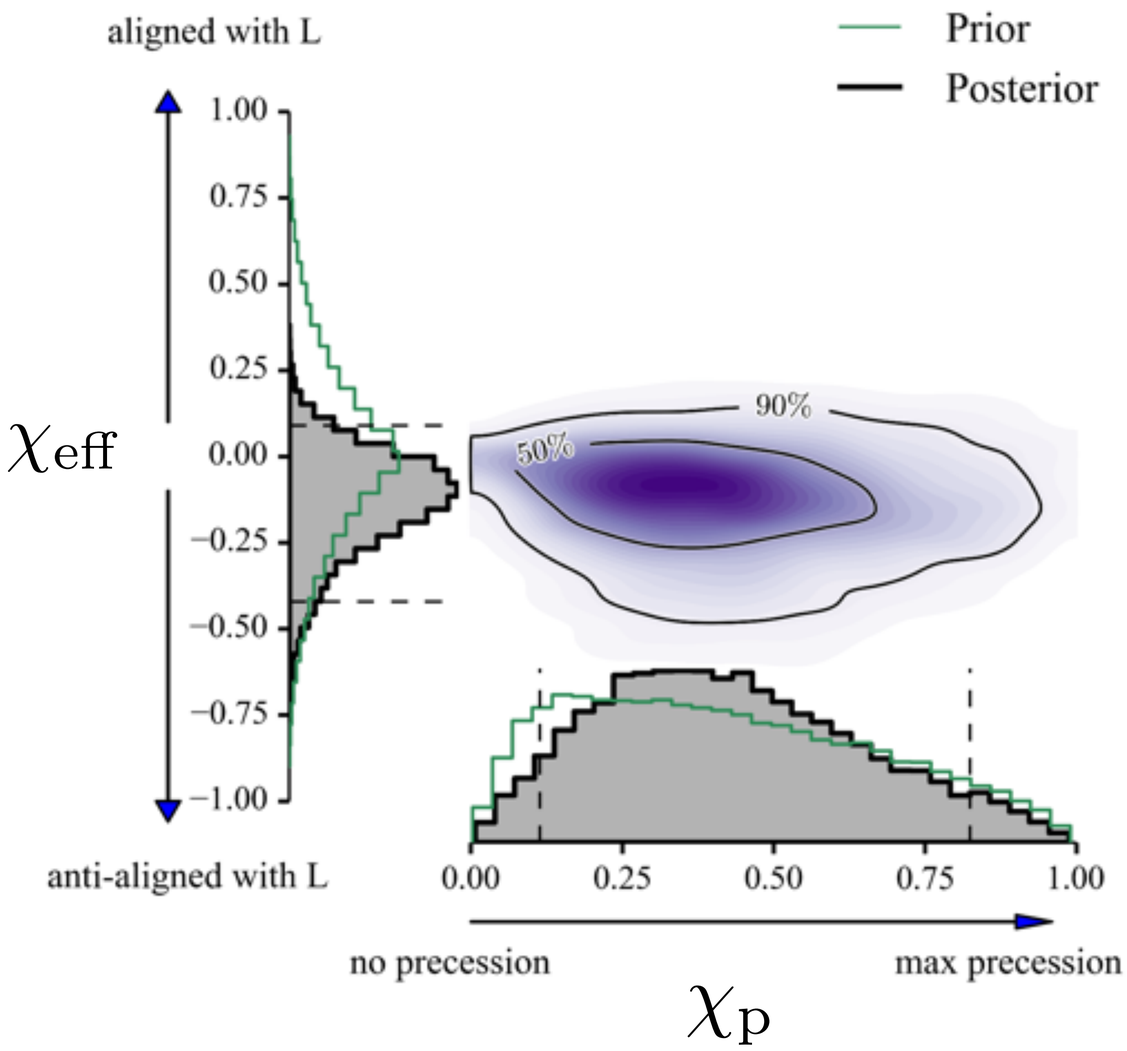
GW170104

BNS GW170817 - Parameter estimation

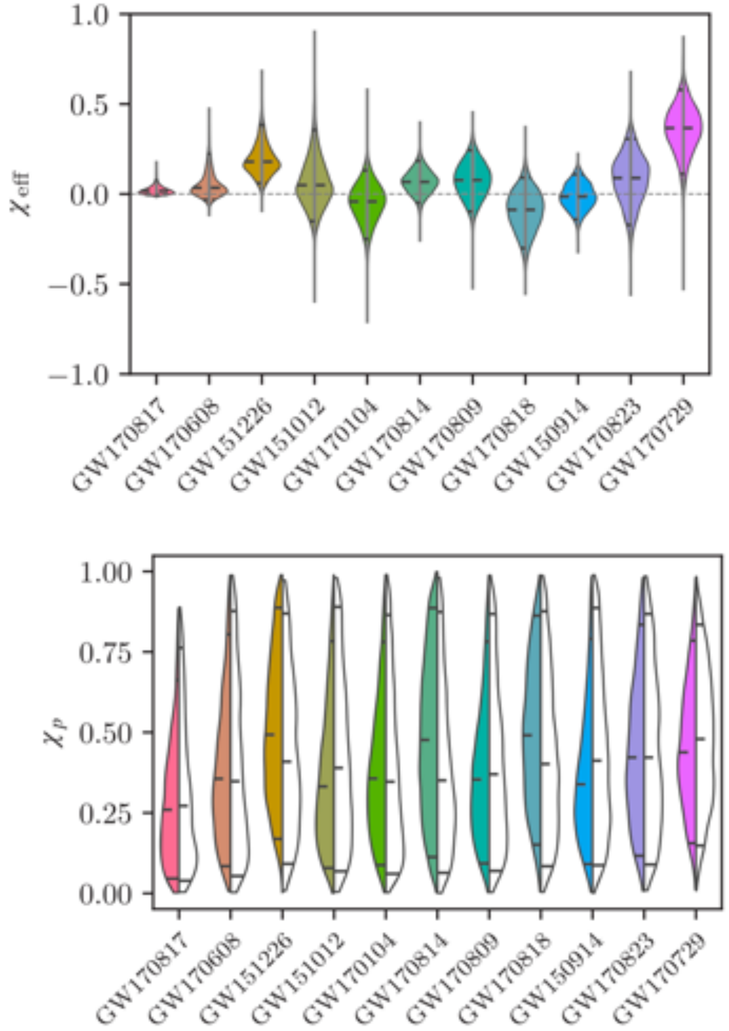


Measuring Black Hole Spins is Hard

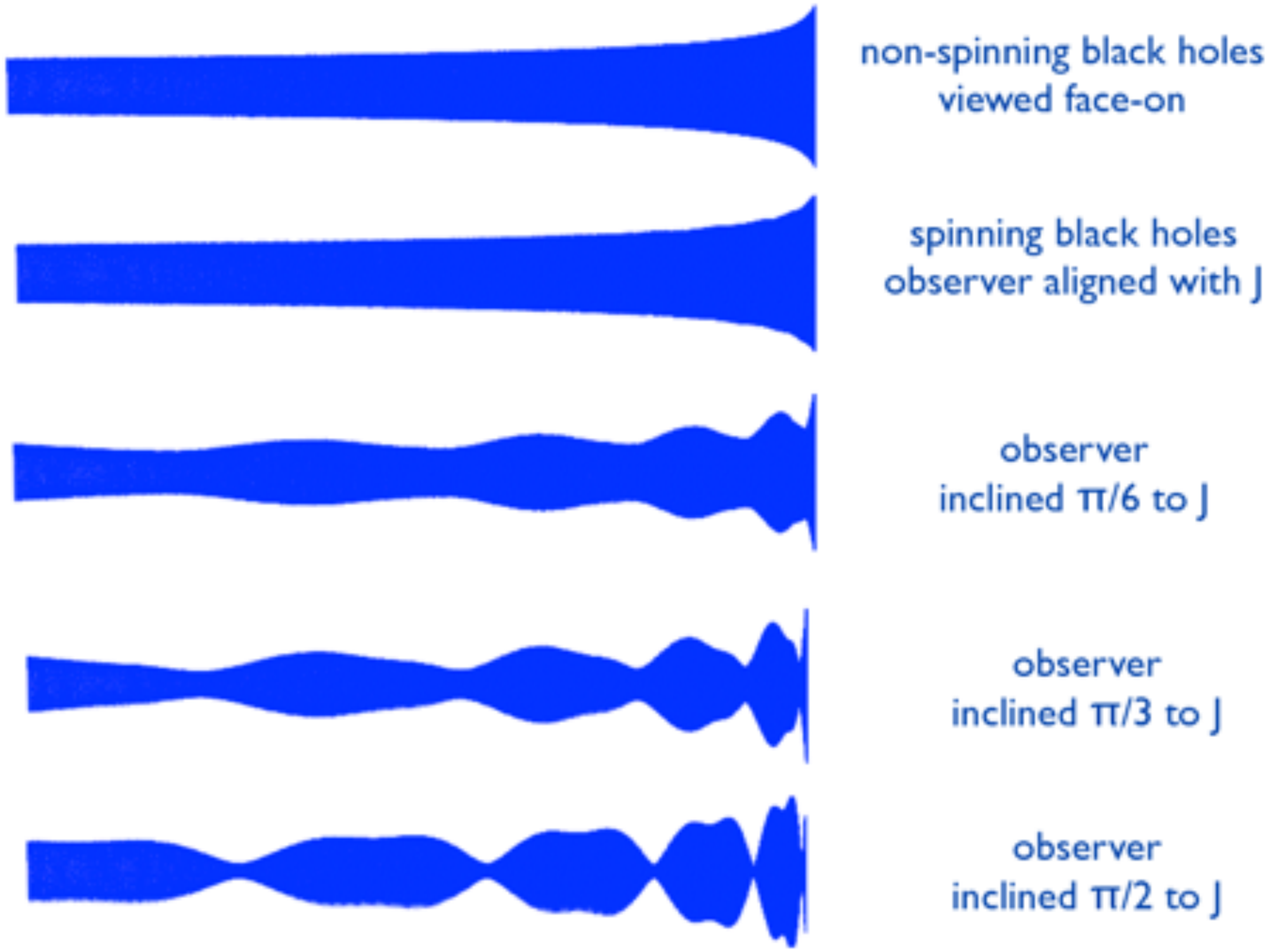
Spin posteriors for GW170104



Spin posteriors GWTC-1



Why measuring BH spin is hard



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: Binary Systems

$$\rho^2 \sim \frac{\mathcal{M}^{5/3}}{D_L^2}\left(1 + 6\cos^2\iota + \cos^4\iota\right)$$

- More sensitive to nearby sources
- More sensitive to high mass systems
- More sensitive to face on/off systems

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

$\Rightarrow$ Harder to measure precession

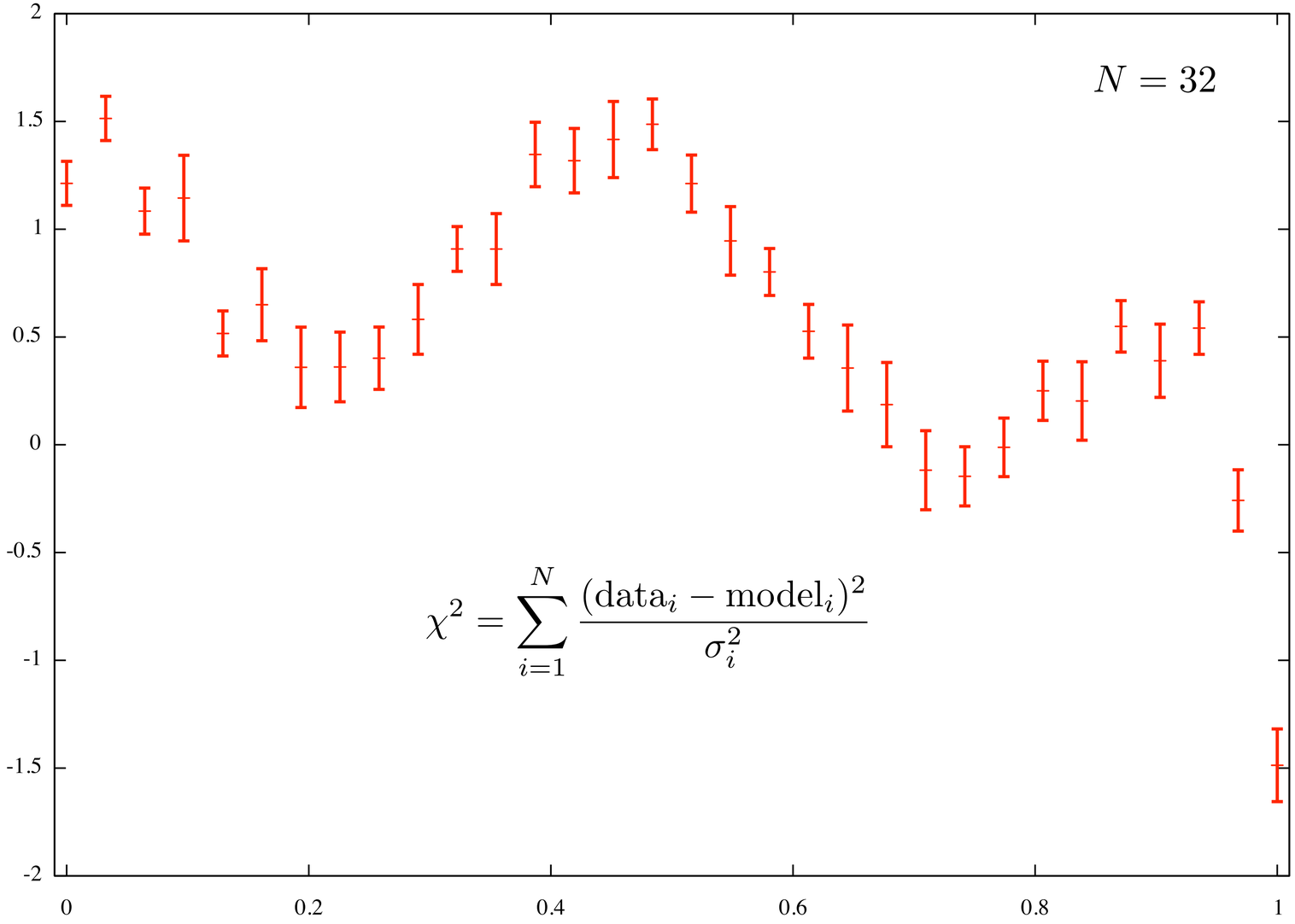
Advanced Techniques: Trans-dimensional Inference

- LISA data analysis - unknown number of signals with unknown parameters
- Unmodelled GW burst - collection of wavelets
- Noise transients, power spectra

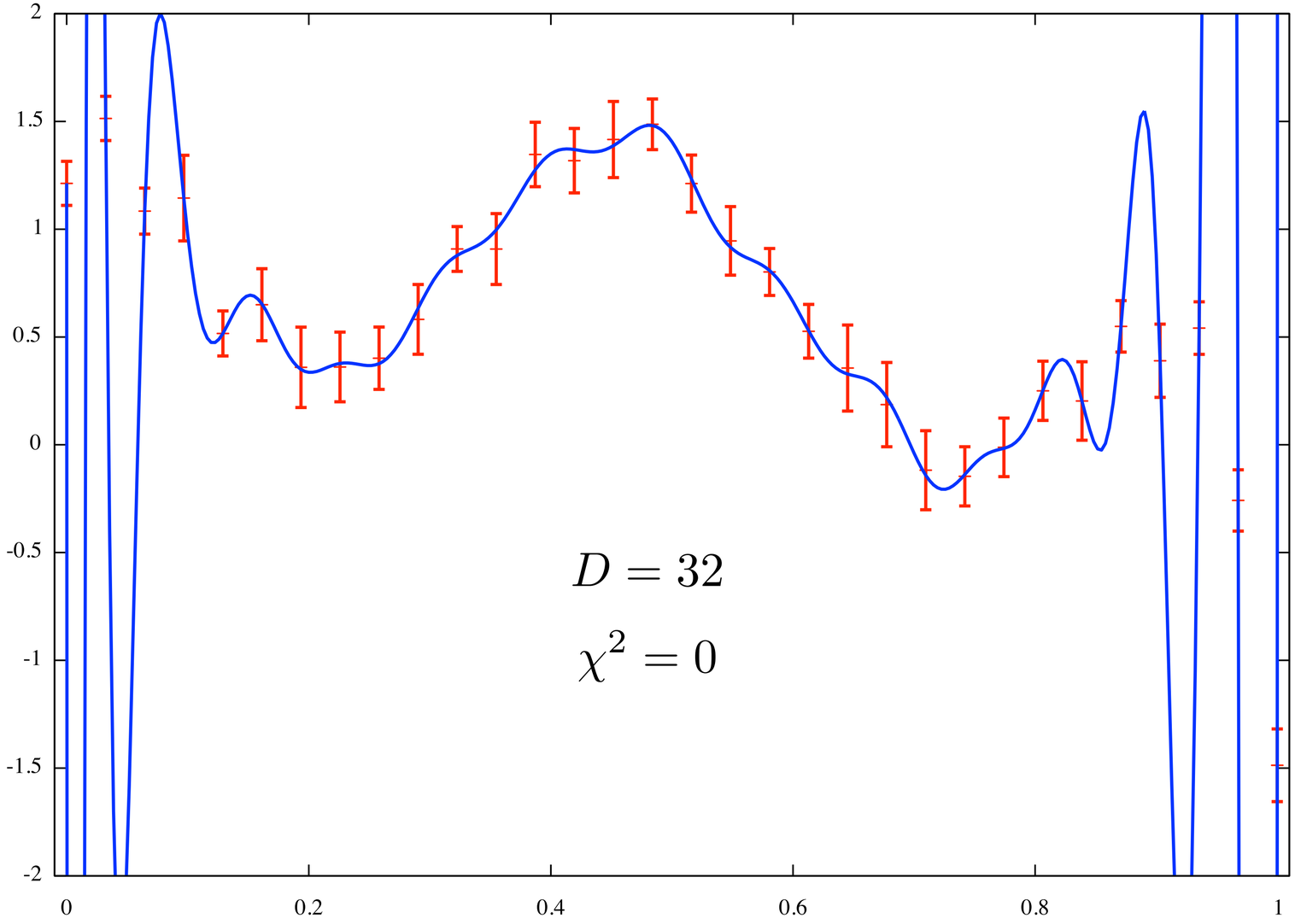
$\Rightarrow$ **Let the data decide the model dimension**

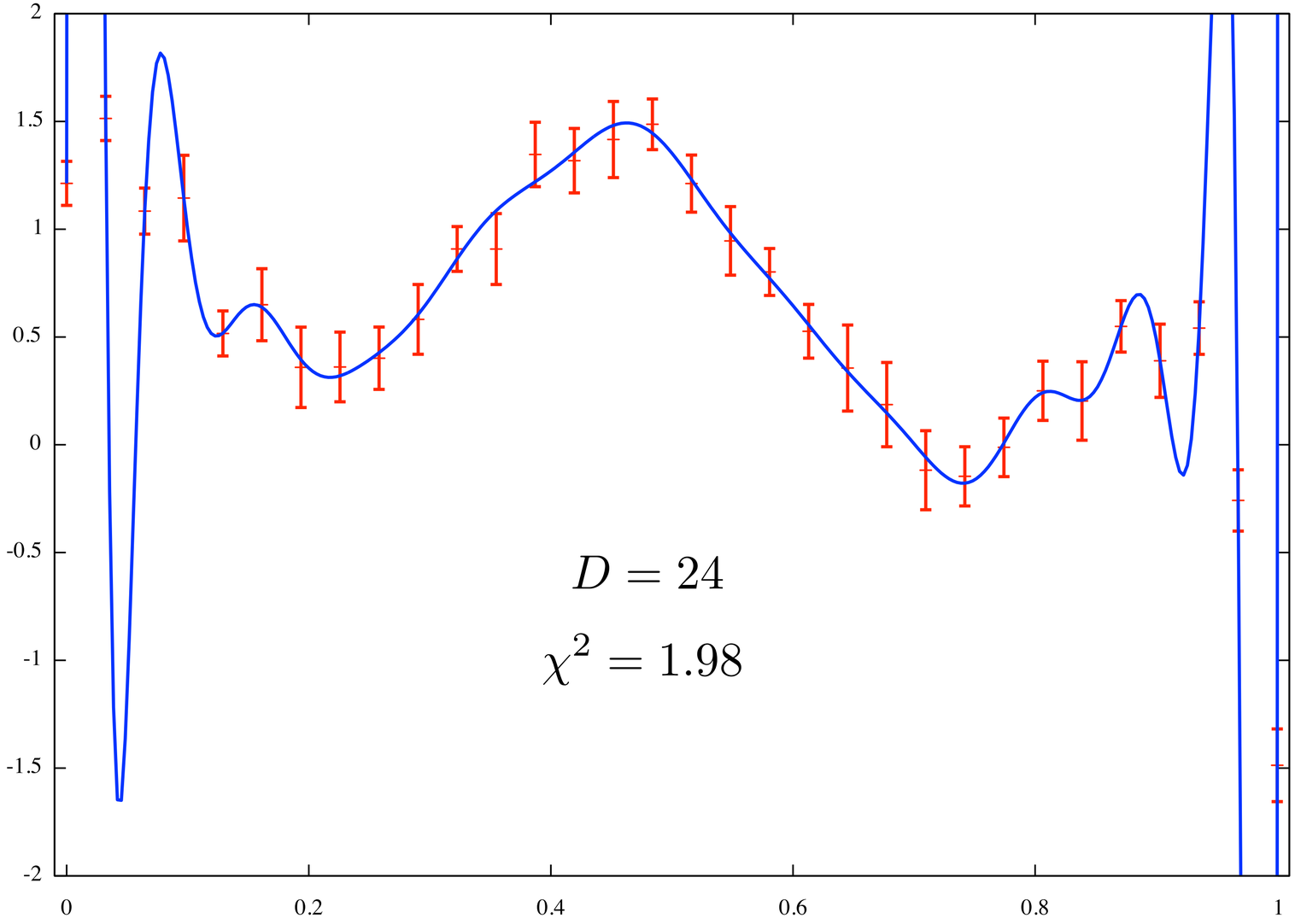
$\Rightarrow$ **Make the model dimension a parameter**

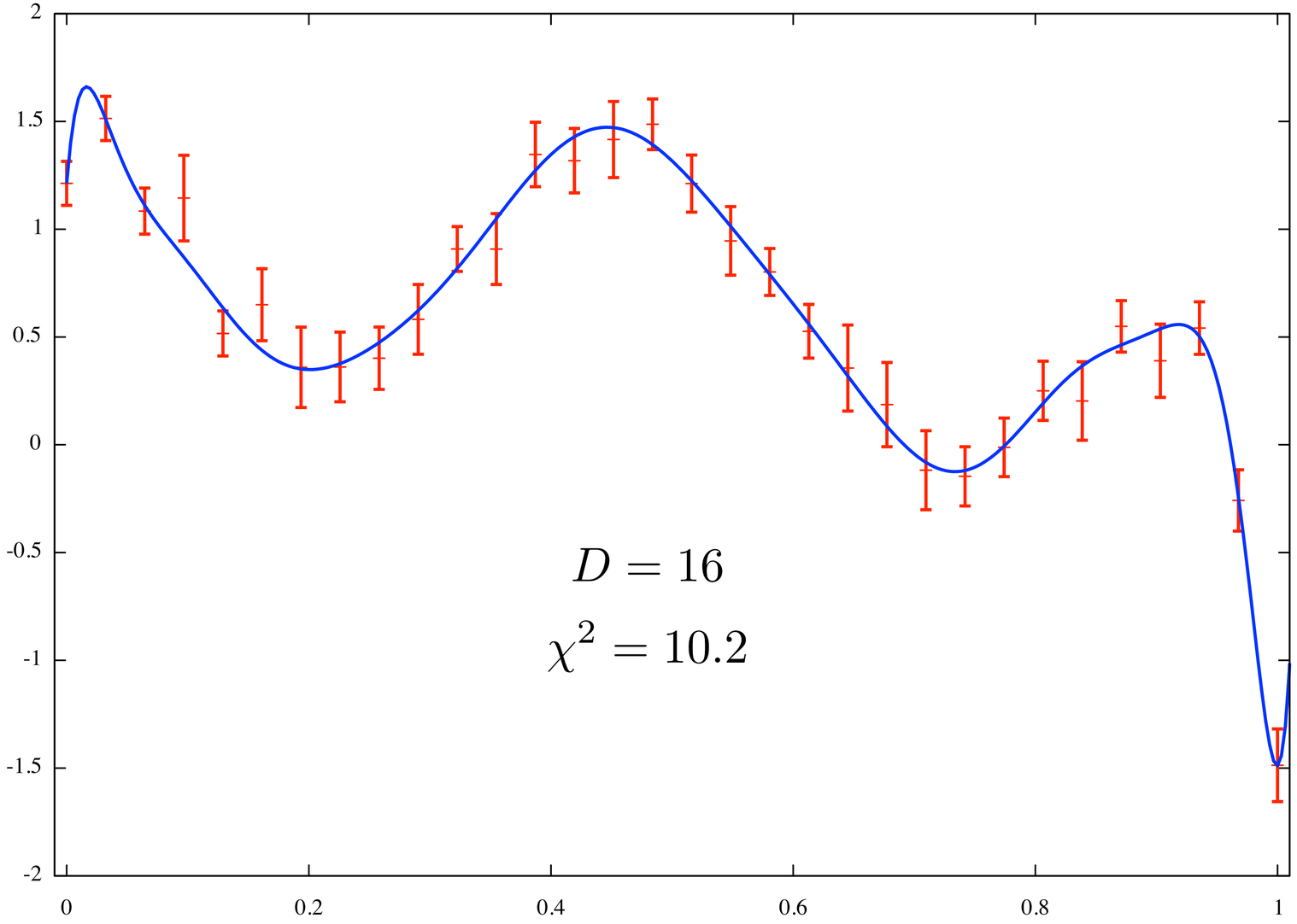
Advanced Techniques: Trans-dimensional Inference

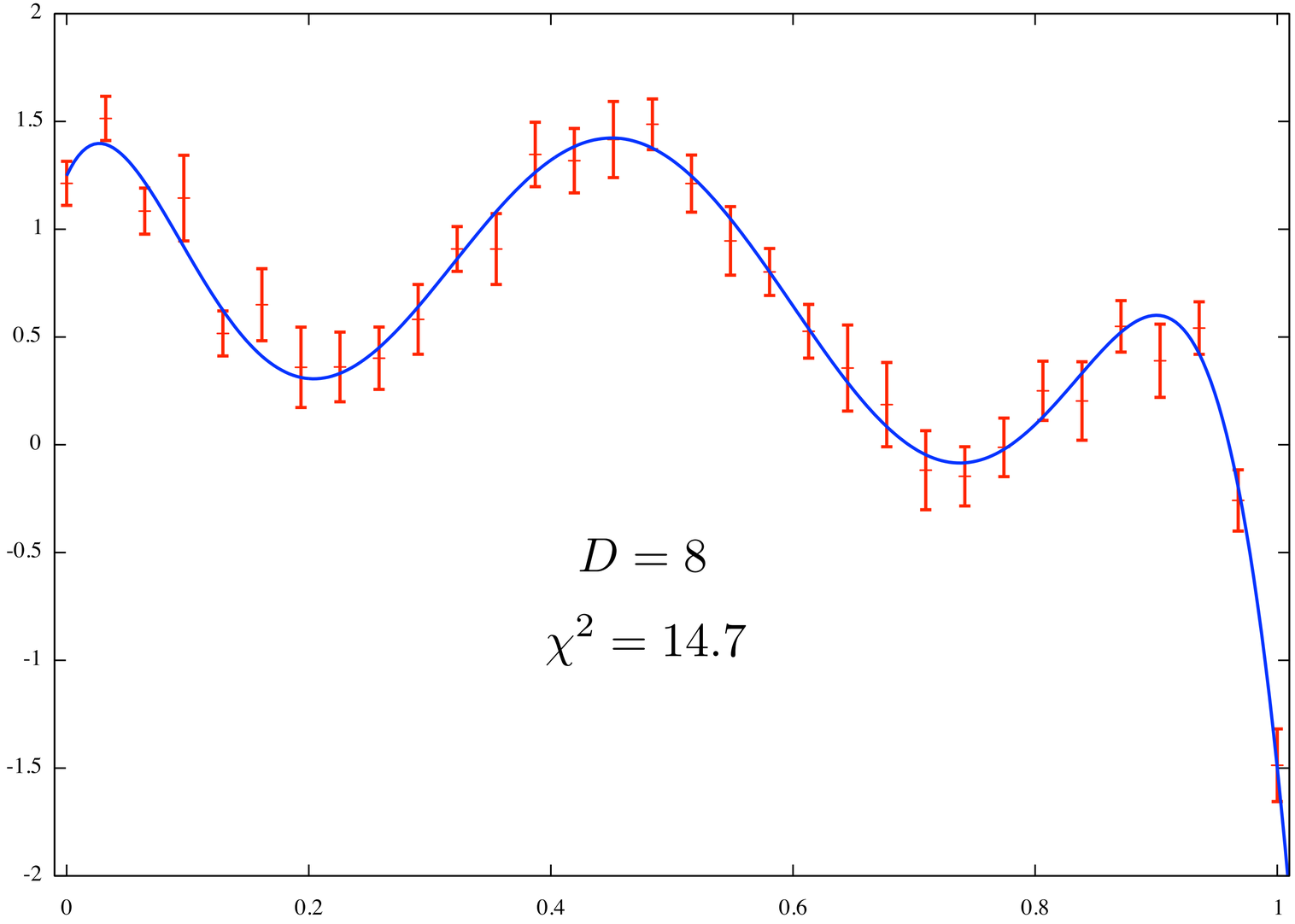


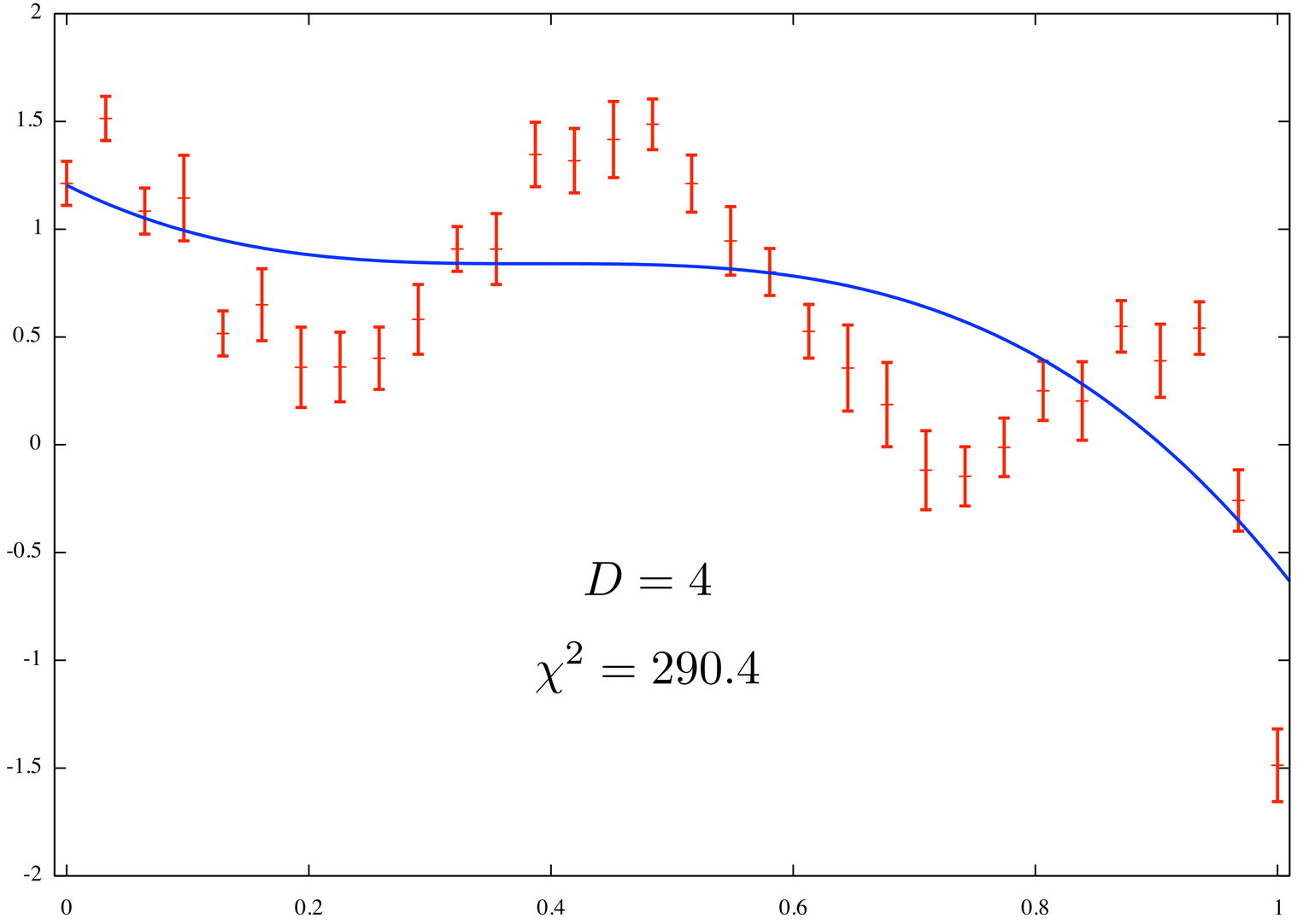
Example, fitting an order D polynomial to N data points


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
$D = 32$
$\chi^2 = 0$

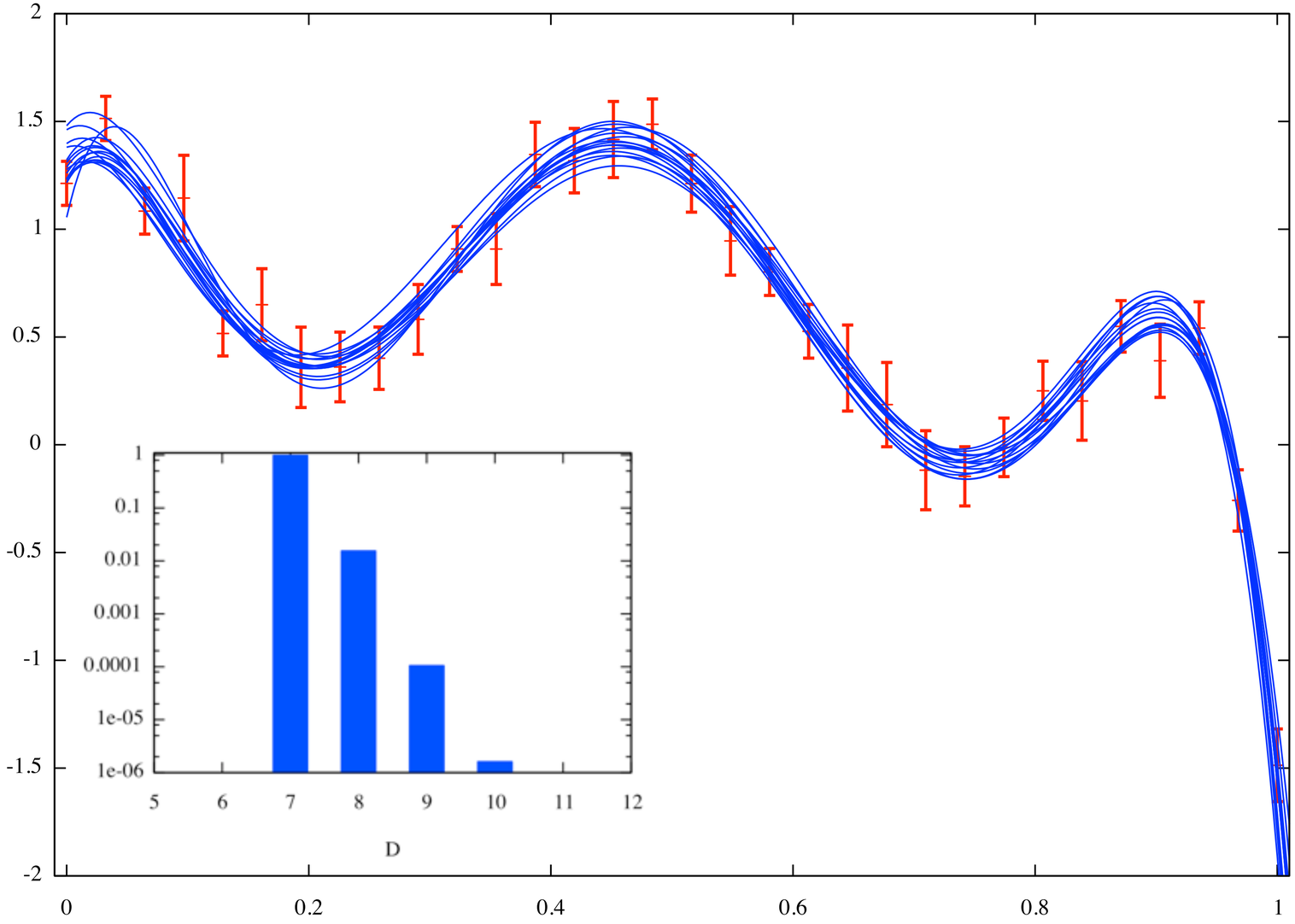

2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
$D = 24$
$\chi^2 = 1.98$


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
$D = 16$
$\chi^2 = 10.2$


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
$D = 8$
$\chi^2 = 14.7$


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
$D = 4$
$\chi^2 = 290.4$

Trans-dimensional Markov Chain Monte Carlo

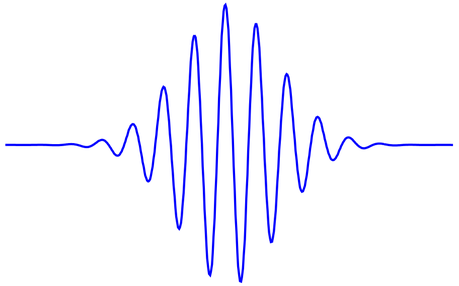


Detection without templates

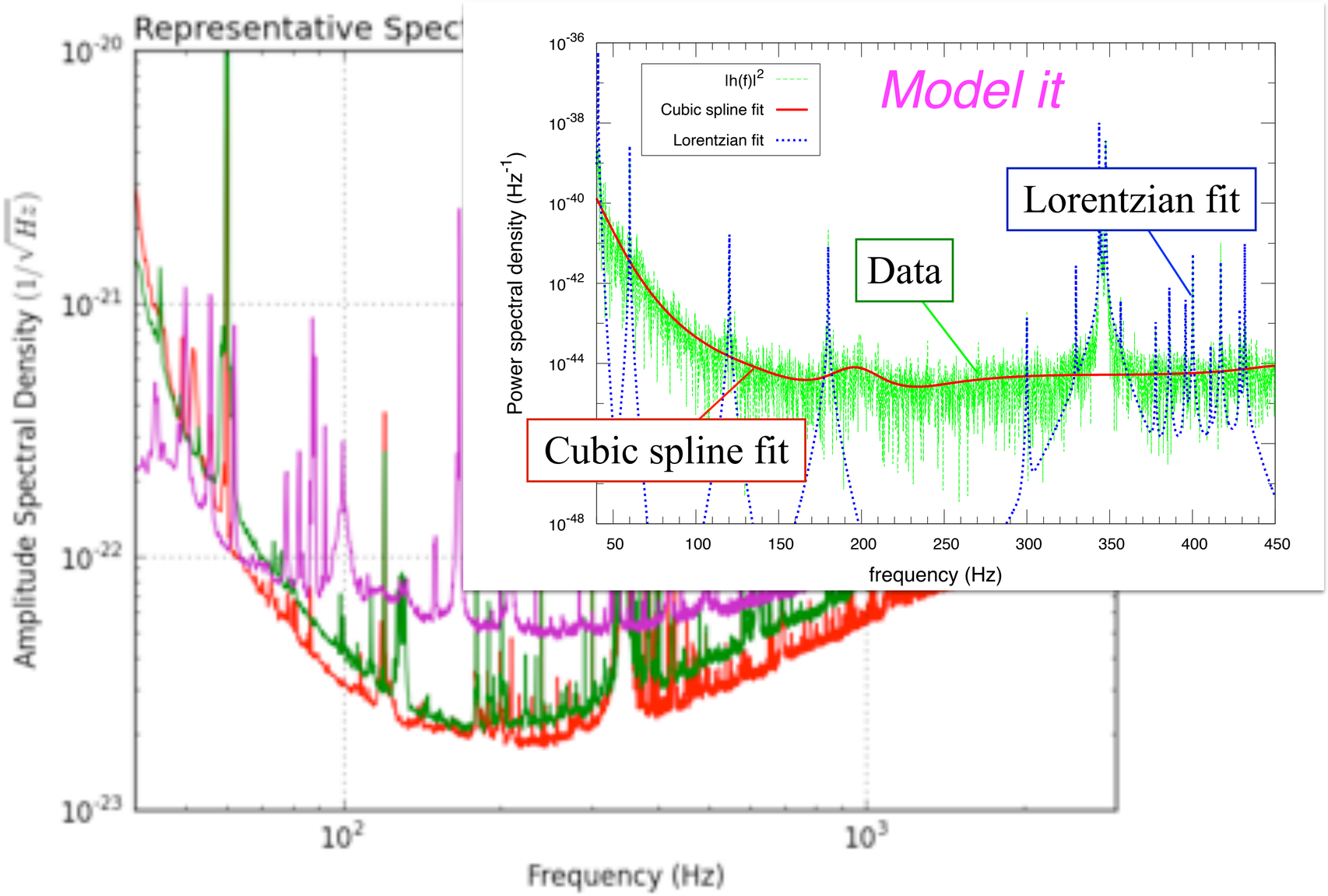
BayesWave

- Bayesian model selection
  - Three part model (signal, glitches, gaussian noise)
  - Trans-dimensional Markov Chain Monte Carlo
- Wavelet decomposition
  - Glitch & GW modeled by wavelets
  - Number, amplitude, quality and TF location of wavelets varies

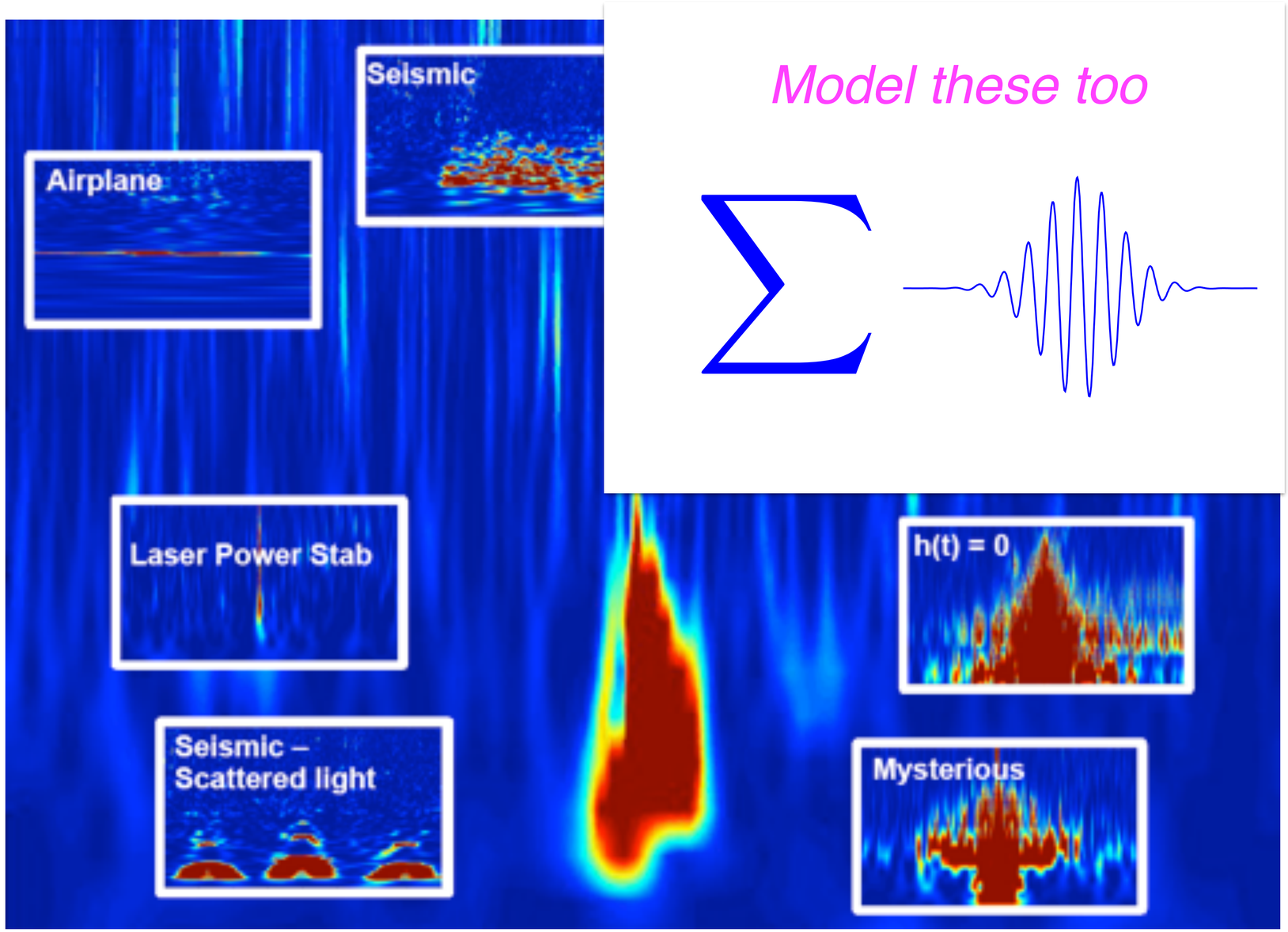
Continuous Morlet/Gabor Wavelets



Lines and a drifting noise floor



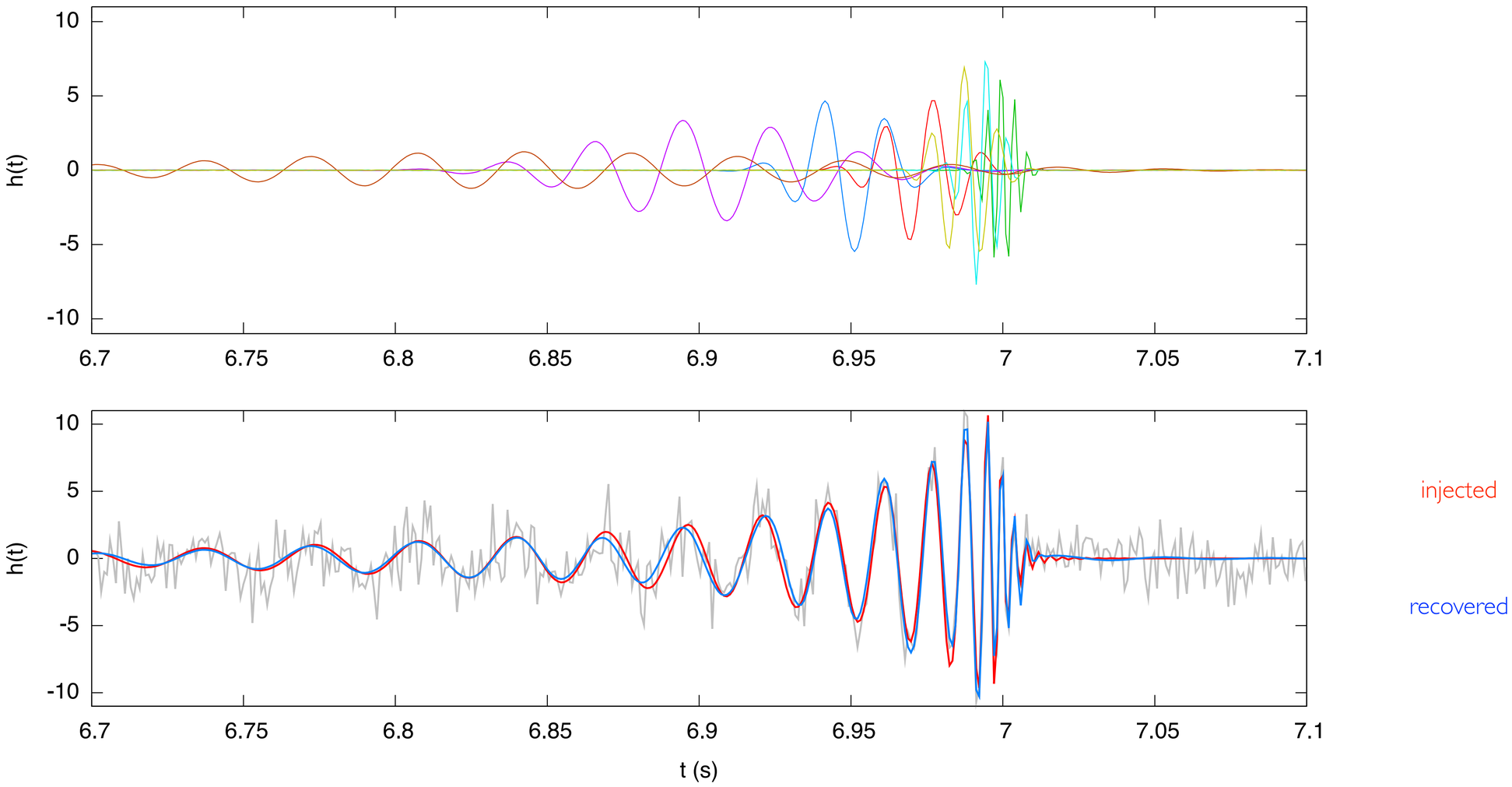
Glitches



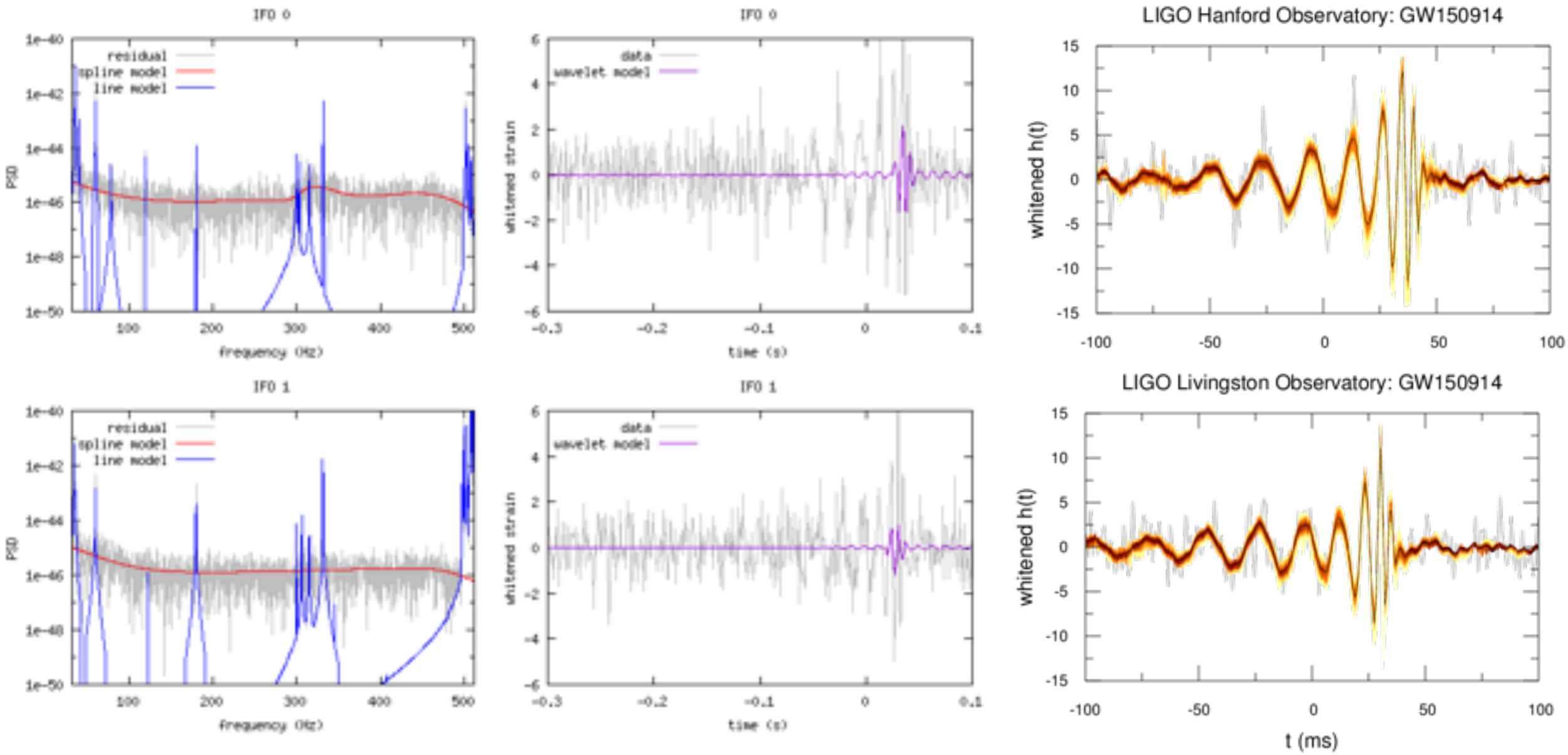
Model these too

Gravitational Waves

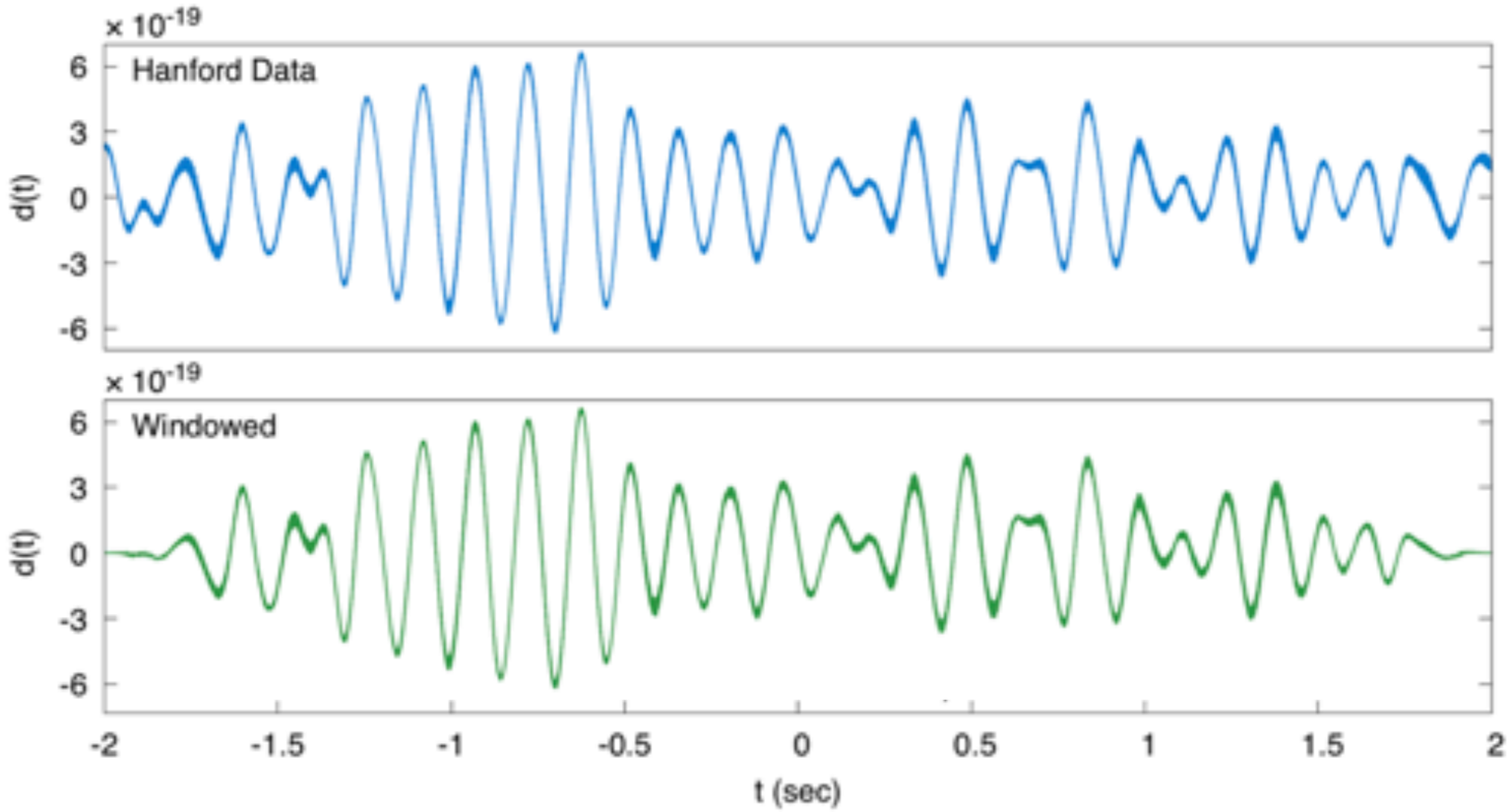
$$p(\mathbf{h}) = \delta(\mathbf{h} - \sum \quad)p(\quad)$$



Reconstructing GW150914 with wavelets



Aside: Power Spectral Estimation

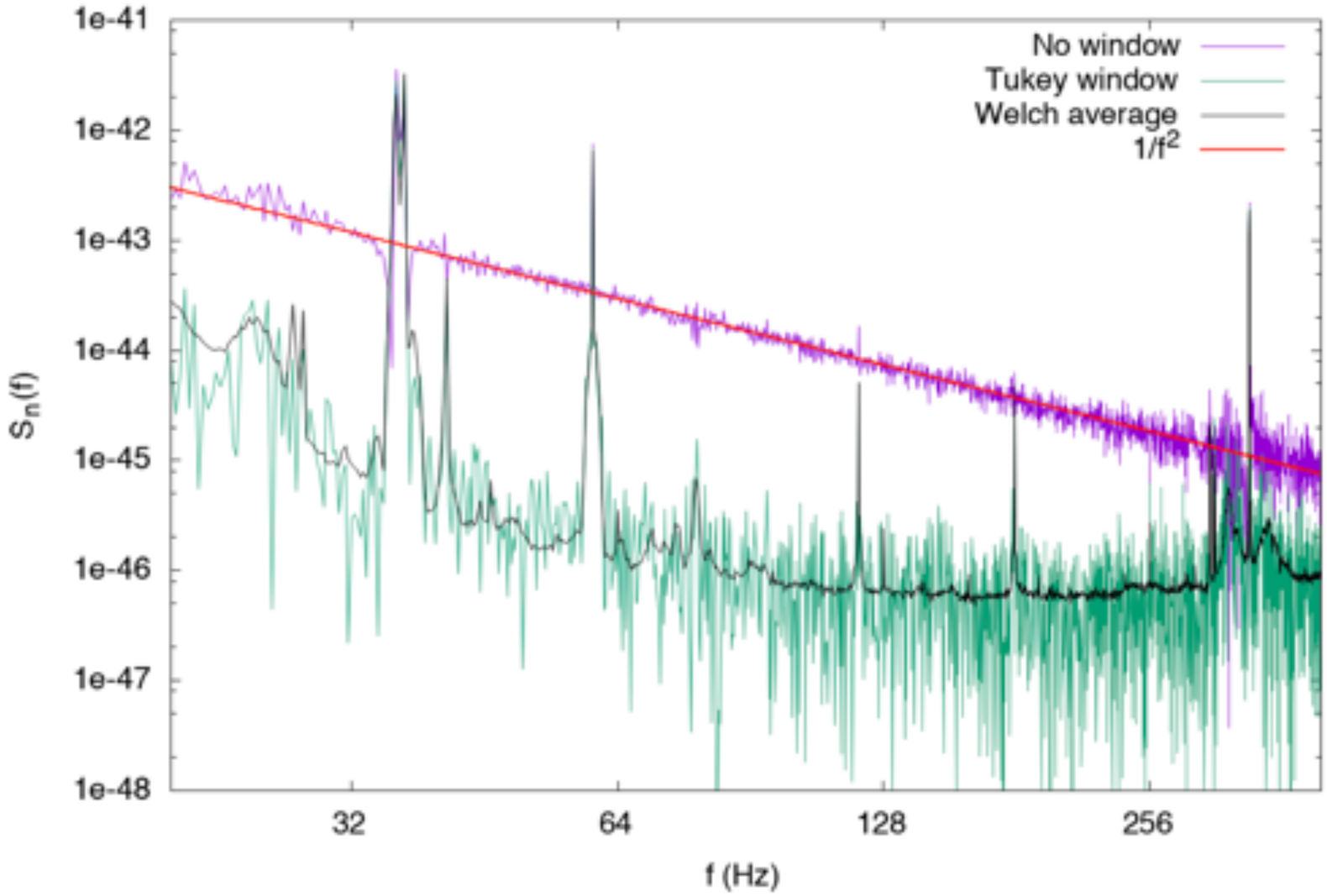


Tukey Window Function

$$w(t) = \begin{cases} \frac{1}{2}\left(1 - \cos(\pi t/\tau)\right) & \text{if } t \leq \tau \\ 1 & \text{if }, \tau < t < T - \tau \\ \frac{1}{2}\left(1 - \cos(\pi(T-t)/\tau)\right) & \text{if } t \geq T - \tau \end{cases}$$

First step is to Fourier transform the data - need to window the data

Aside: Power Spectral Estimation



Welch averaging breaks the data up into chunks and averages the periodograms from each segment. Only valid if the data is stationary.

A Periodogram is the squared magnitude of the Fourier coefficients

$$P_i = |\tilde{d}_i|^2$$

The green line here is a periodogram of the 4 seconds of data shown on the previous slide

Aside: Power Spectral Estimation

More advanced methods fit models using the Whittle likelihood function

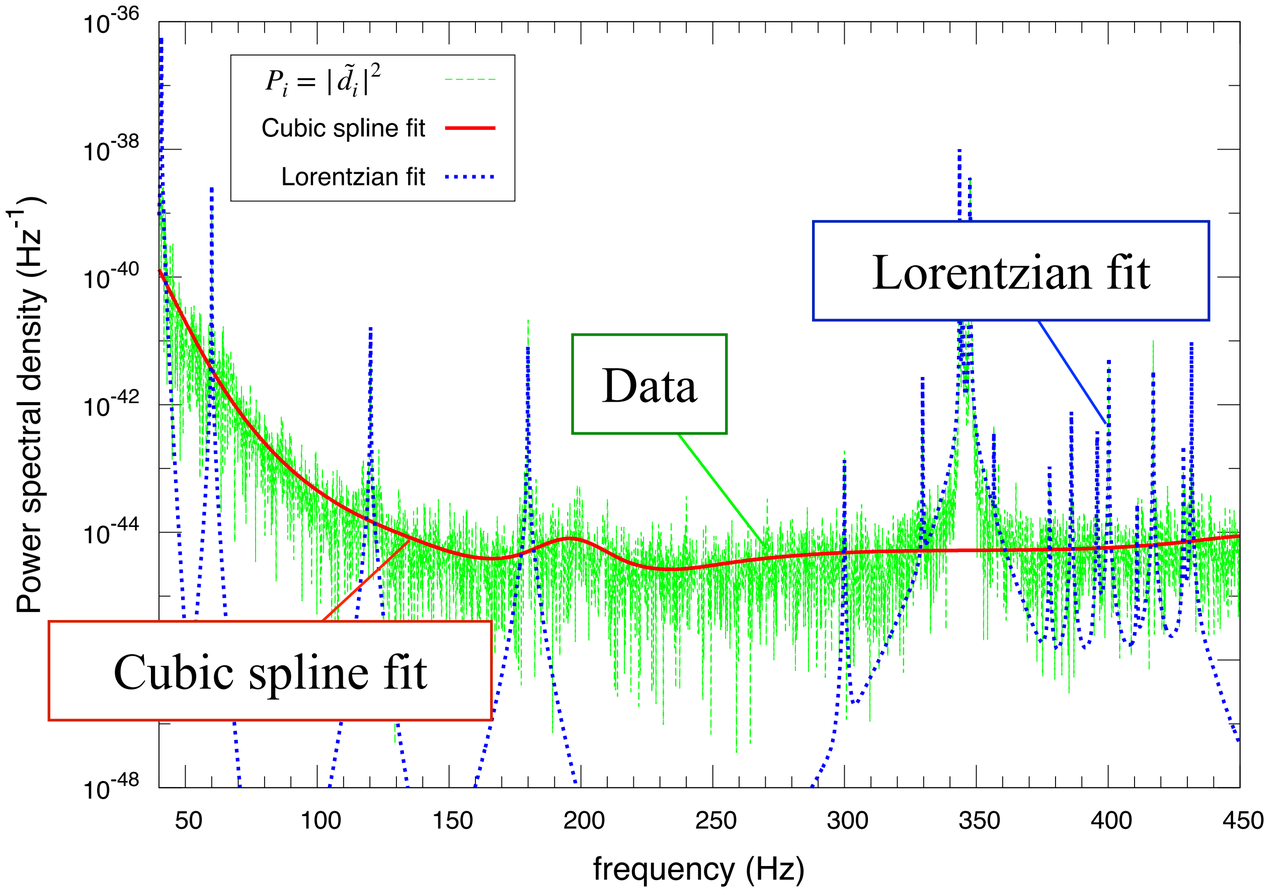
$$\ln p(d \mid S) = -\frac{1}{2} \sum_i \left(\ln S(f_i) + \frac{|\tilde{d}_i|^2}{S(f_i)} \right)$$

For example, we can fit some functional form to the power spectral density. E.g. $S(f) = f^{\alpha}$

Aside: Power Spectral Estimation

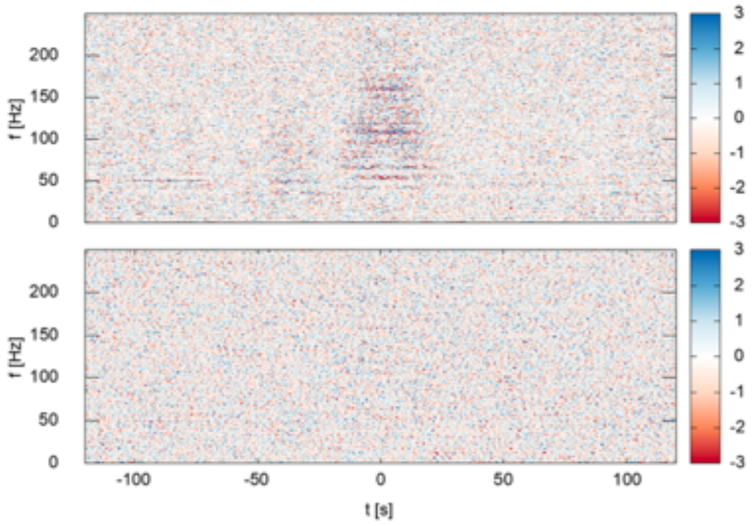
$$\ln p(d\,|\,S) = -\frac{1}{2}\sum_i \left(\ln S(f_i) + \frac{|\tilde{d}_i|^2}{S(f_i)} \right)$$

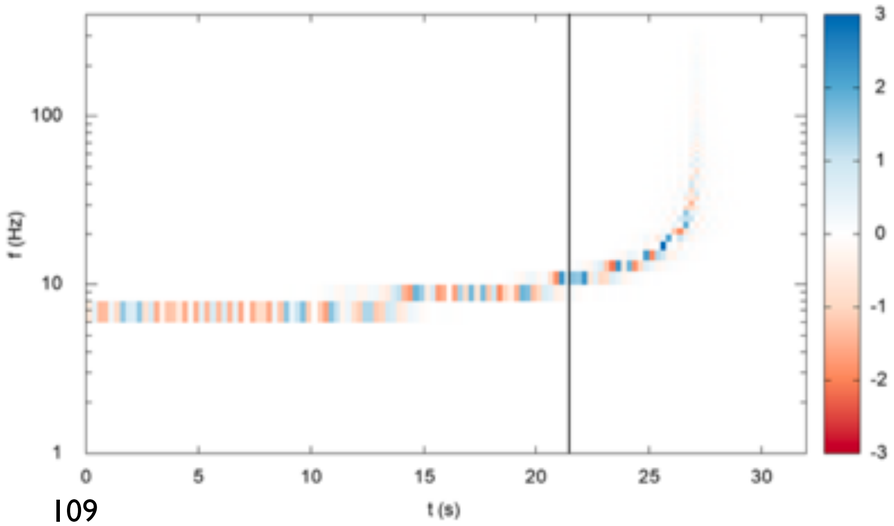
BayesWave models the spectrum using a smooth cubic spline plus Lorentzian lines



Challenges in GW data analysis

- Searching for precessing/eccentric signals (high dimension)
- LISA - complicated instrument response, thousands of overlapping signals
- Non-stationary and non-Gaussian noise





Non-stationary 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

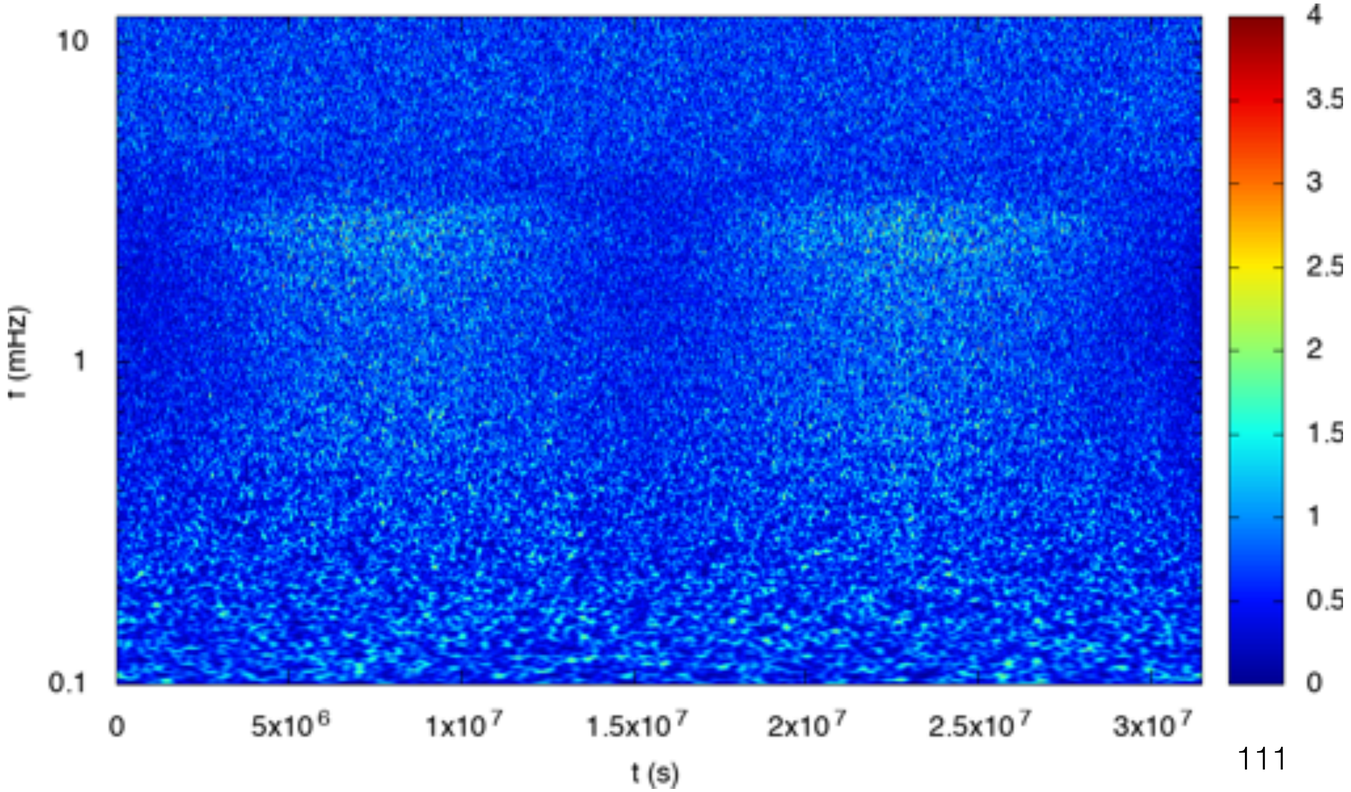
1. [“Fitting time series models to nonstationary processes”. Dahlhaus, Ann. Statist., 25, 1 (1997)]

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

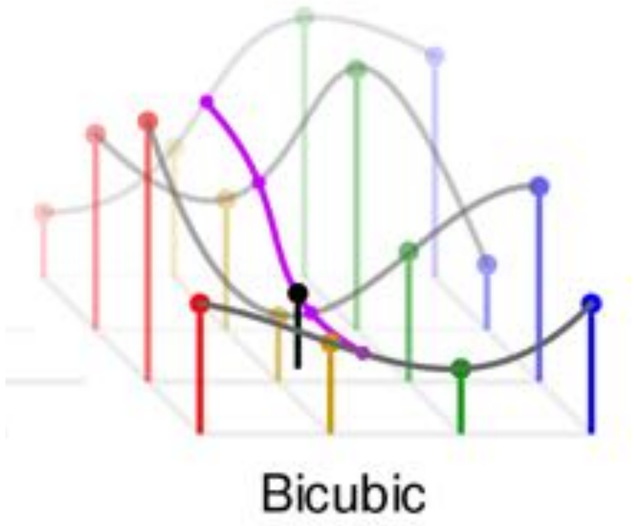
Non-stationary Noise

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

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

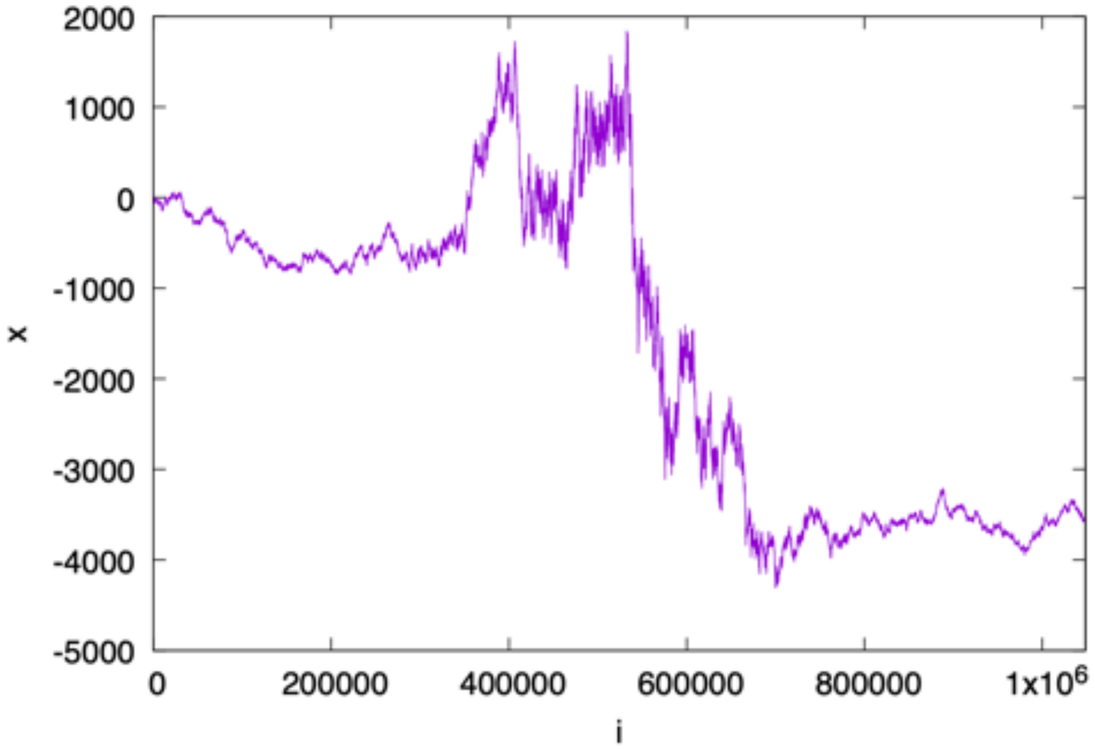


Model the wavelet spectrum C_{ik} as a smooth function in frequency and time. E.g. Trans-dimensional Bicubic spline

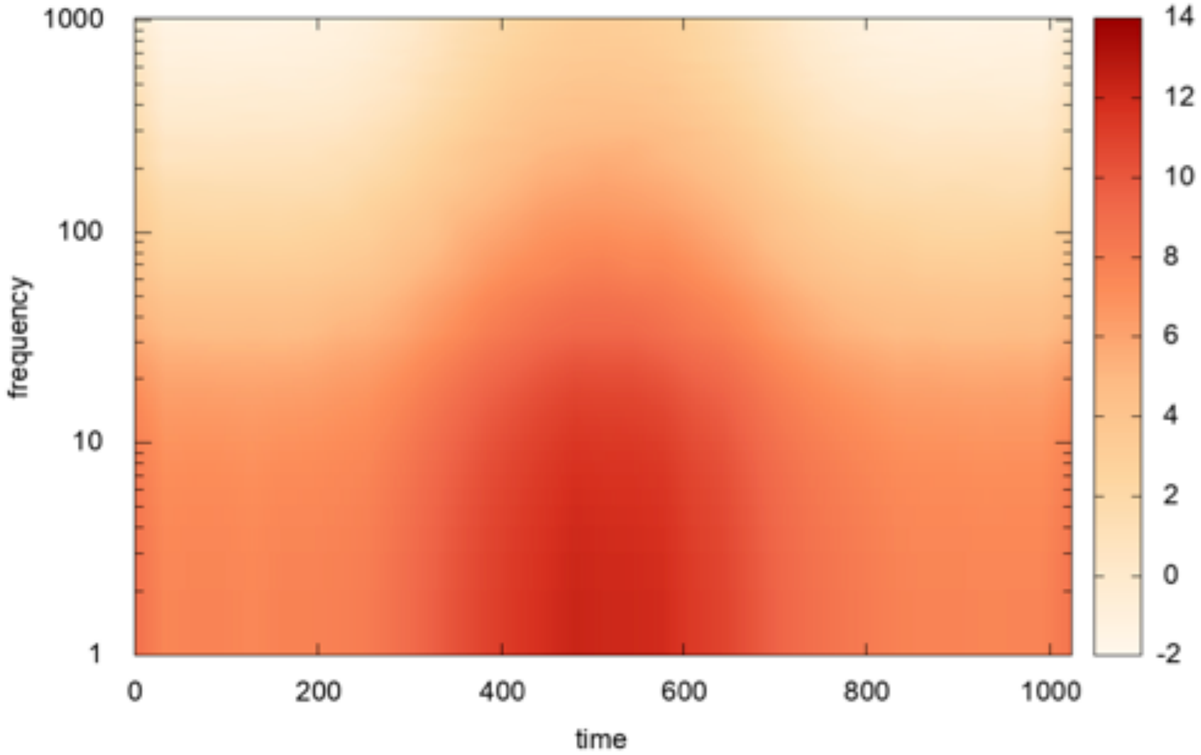


[Cornish, Phys Rev D **102,** 124038 (2020)]

Non-stationary Noise



Non-stationary time series

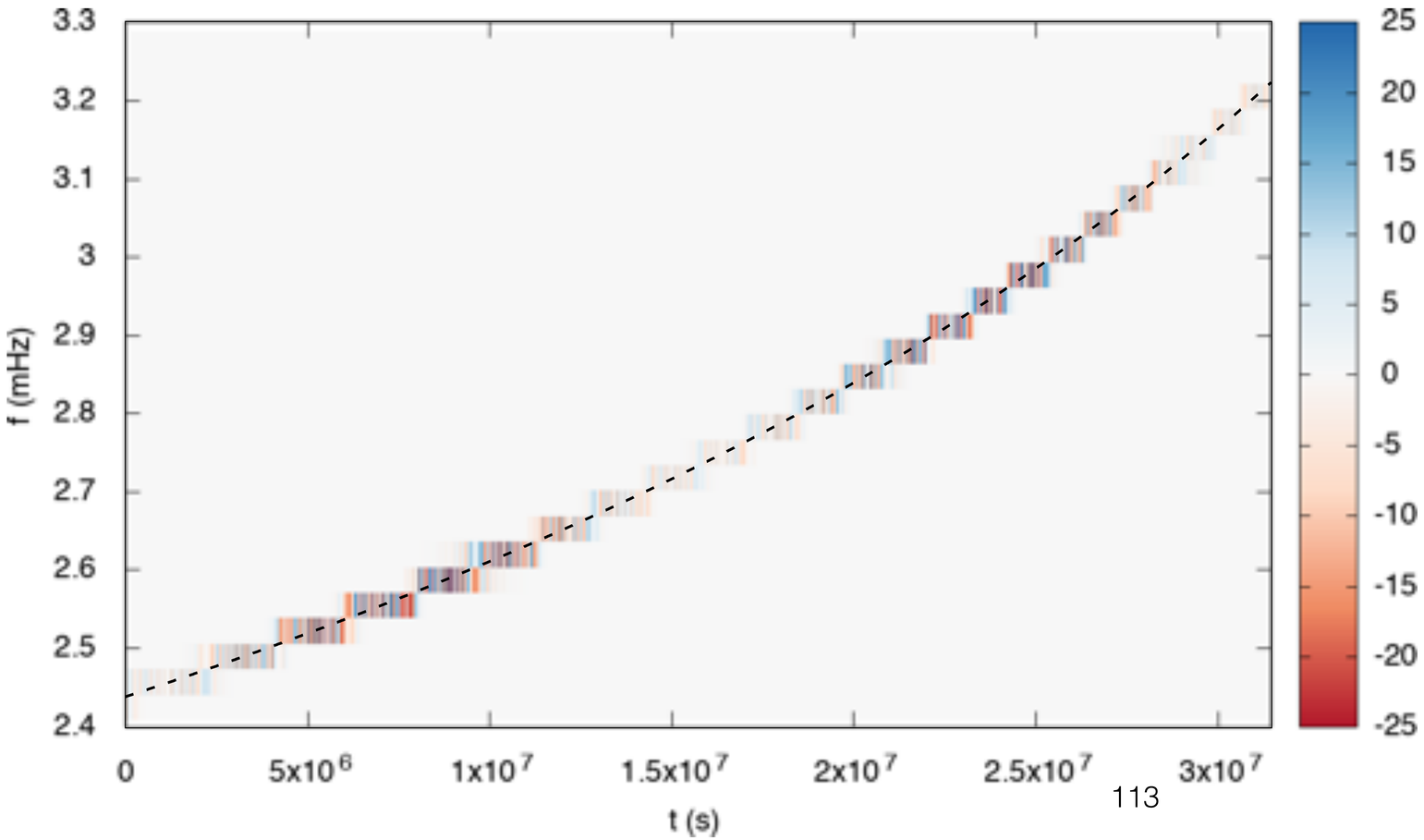


Bi-cubic spline fit to the dynamic spectrum

[Cornish, Phys Rev D **102,** 124038 (2020)]

Wavelet domain waveforms

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

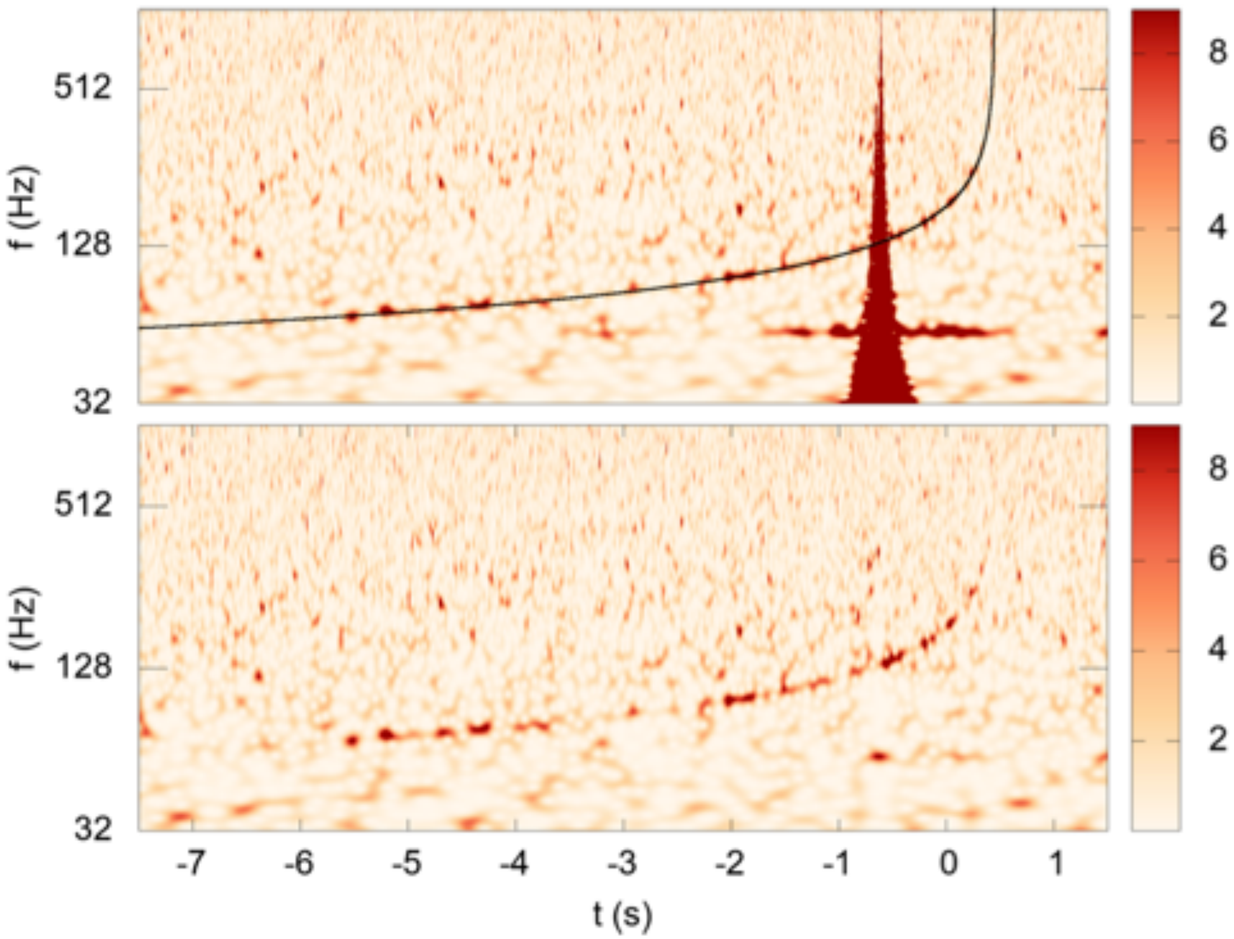


Fast wavelet transforms of the signals for computational efficiency

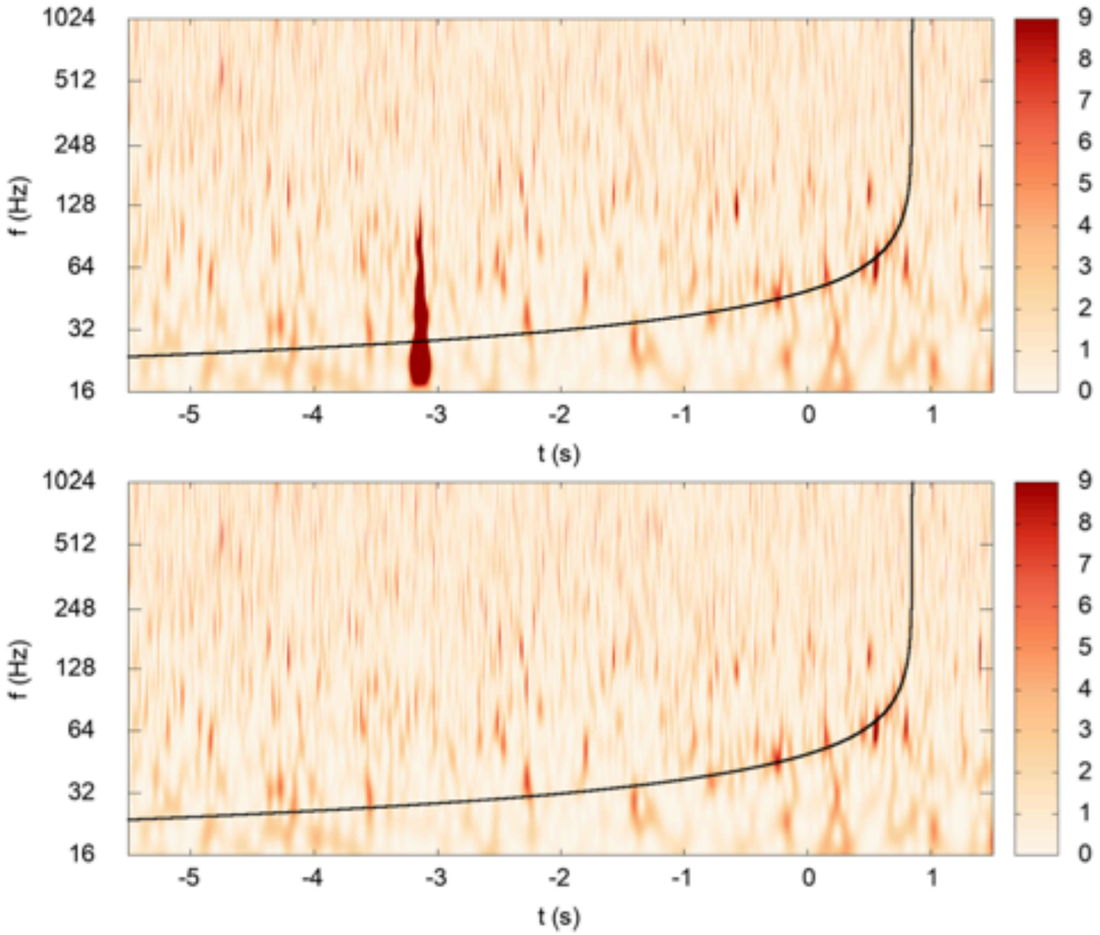
Faster than frequency domain, $\sqrt{N}$ scaling

[Cornish, Phys Rev D **102,** 124038 (2020)]

Noise transients (glitches) and parameter estimation



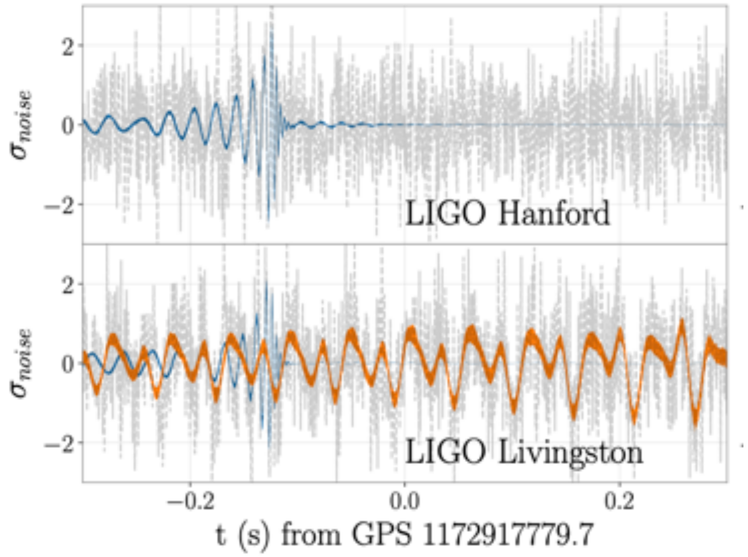
GW170817 (Livingston)



GW190924_021846 (Livingston)

[Cornish, 2101.01188 (2021)]

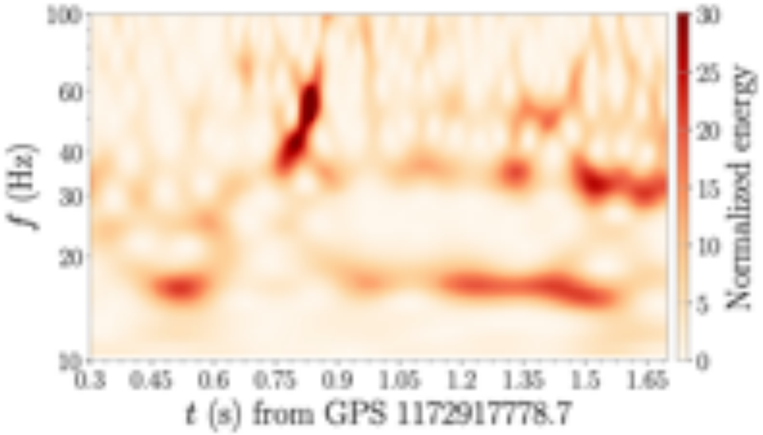
Joint inference of signals and glitches



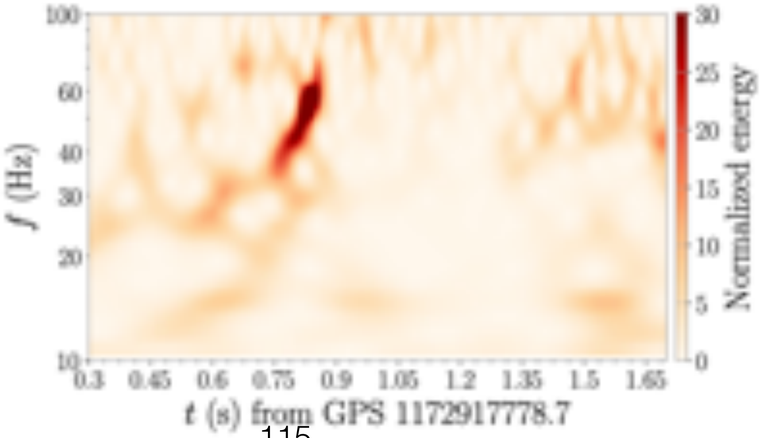
BayesWave can now simultaneously model CBC signals using template, glitches using wavelets and the power spectrum using splines and lines

[Chatziioannou+, Phys. Rev. D **103** 044013 (2021)]

Livingston Data



Livingston Data - Glitch



Livingston Data - Glitch - Signal