Image: LIGO / Caltech / MIT / Sonoma State (Aurore Simonnet)

Introduction to Bayesian Parameter Estimation for Compact Binary Coalescences

Sylvia Biscoveanu

GWATUT

Sept. 20th 2021

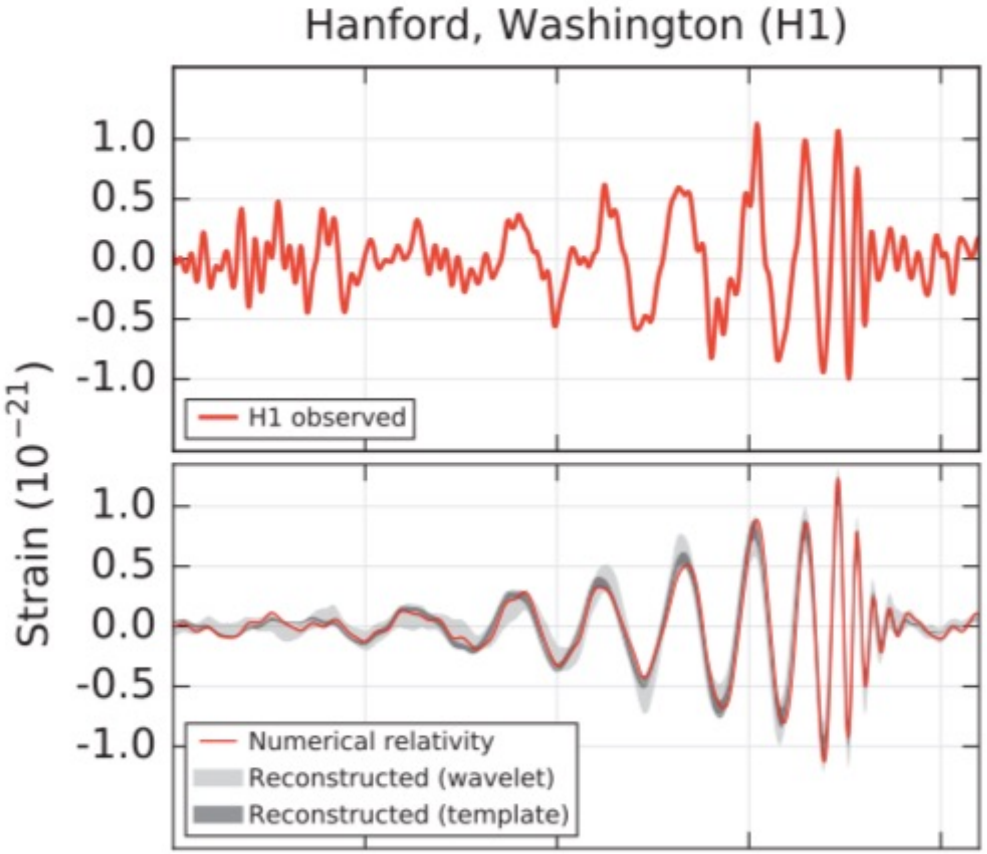
MIT Kavli Institute for Astrophysics and Space Research

Overview

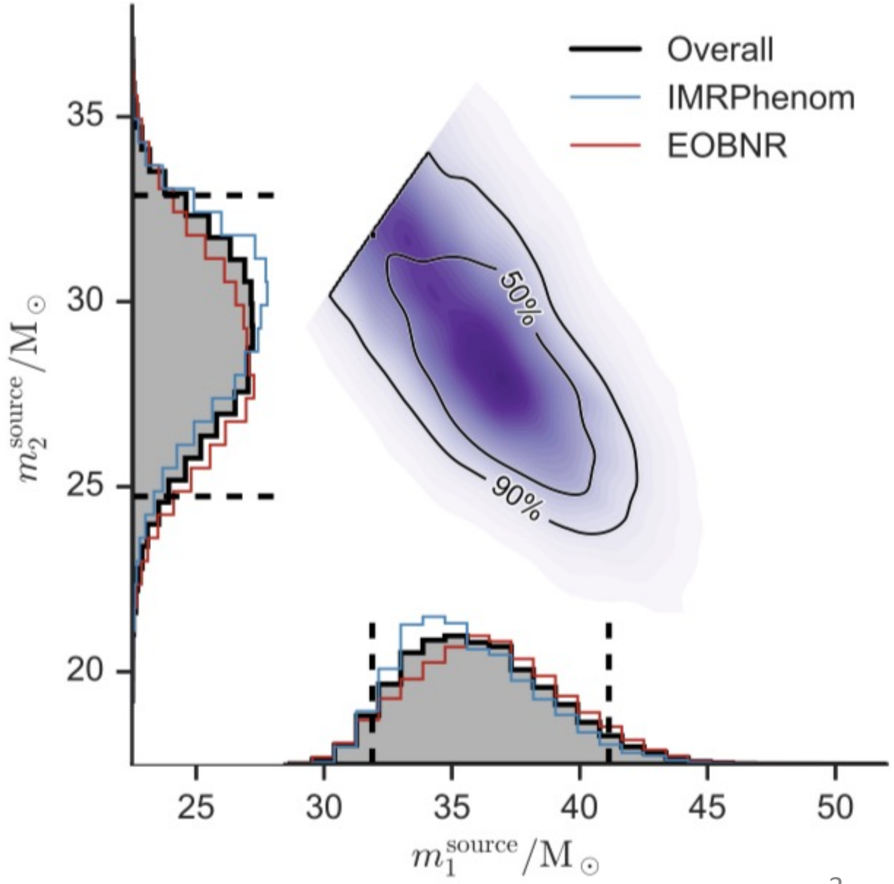
- Motivation
- Statistical Introduction
- LIGO Noise
- Gravitational-wave likelihood
- Astrophysical parameters and priors
- Model Selection
- Sampling methods
- Example

Motivation

- How do we go from this… to this?



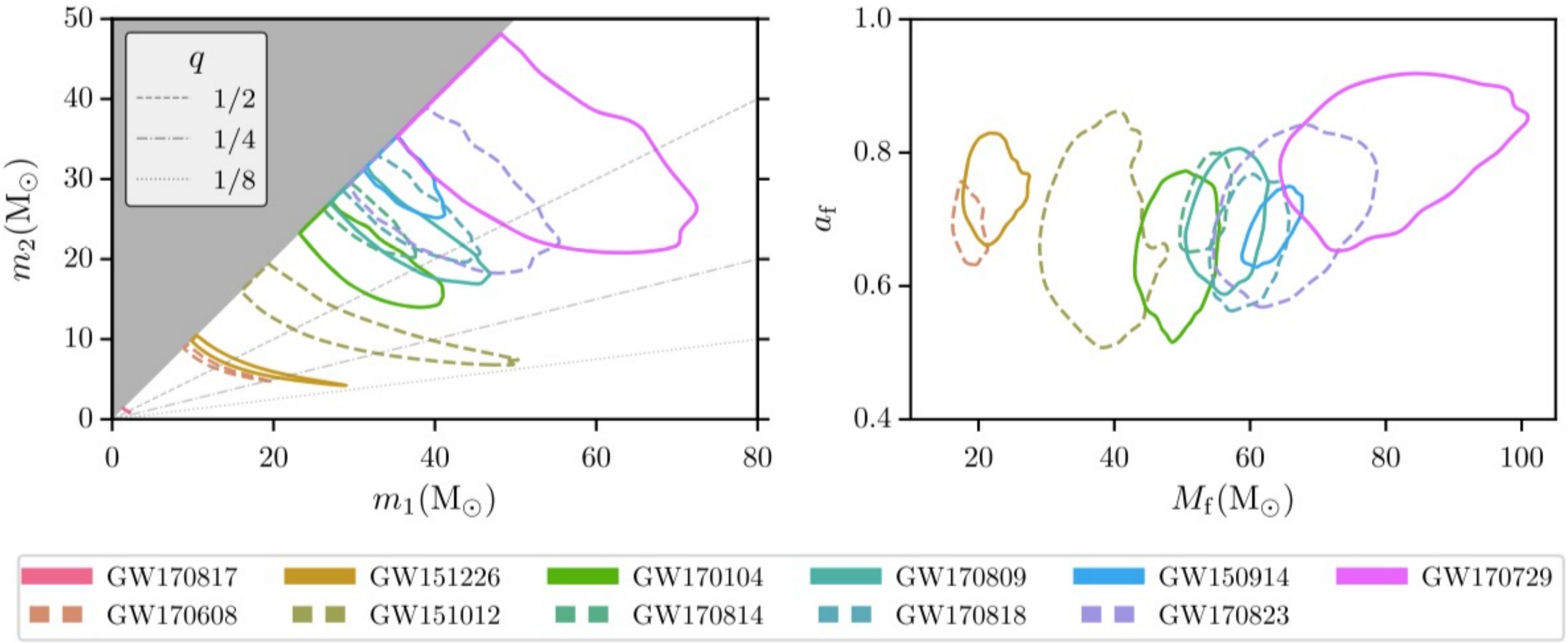
https://arxiv.org/abs/1602.03837



https://arxiv.org/pdf/1602.03840v2.pdf

Motivation

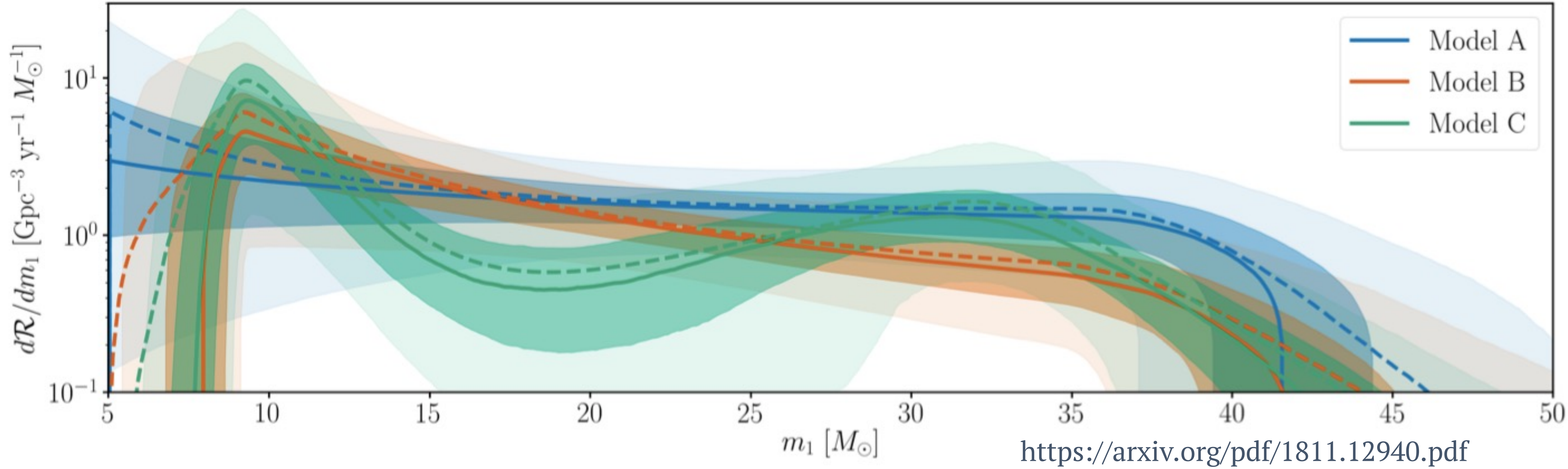
- Measure the properties of individual detections…



https://arxiv.org/pdf/1811.12907.pdf

Motivation

- … and of the underlying population
  - See Salvo Vitale’s talk later today!



https://arxiv.org/pdf/1811.12940.pdf

Bayes' Theorem

$$\overset{\text{posterior}}{p(\underset{\text{parameters}}{\boldsymbol{\theta}}|\underset{\text{data}}{d}, \underset{\text{model}}{H})} = \frac{\overset{\text{likelihood}}{p(d|\boldsymbol{\theta}, H)}\,\overset{\text{prior}}{p(\boldsymbol{\theta}|H)}}{\underset{\text{evidence}}{p(d|H)}}$$

Bayes' Theorem Components

- Posterior – probability of the parameters θ given the data d and model H

$$p(\boldsymbol{\theta}|d, H)$$

- Likelihood – probability of the data d for parameters θ and model H

$$p(d|\boldsymbol{\theta}, H) \equiv \mathcal{L}(d|\boldsymbol{\theta}, H)$$

- Prior – initial probability of the parameters θ under model H

$$p(\boldsymbol{\theta}|H) \equiv \pi(\boldsymbol{\theta}|H)$$

Normalization

- Probability distributions need to be normalized:

$$\int p(\mu)d\mu = 1$$

- Marginalization – certain parameters are integrated out of the distribution

$$\mathcal{L}(d|\mu) = \int \mathcal{L}(d|\theta,\mu)\pi(\theta)d\theta$$

- Evidence – normalization constant for the posterior, likelihood marginalized over all parameters

$$p(d|H) \equiv \mathcal{Z}_H = \int \mathcal{L}(d|\boldsymbol{\theta},H)\pi(\boldsymbol{\theta}|H)\ d\boldsymbol{\theta}$$

Joint Distributions

- Sometimes, the distribution of one parameter can depend on another parameter:

$$p(\theta, \mu) = p(\theta|\mu)p(\mu)$$

  - Example: The secondary mass in the binary must be less than the primary mass

- In this case, marginalization can look like:

$$\mathcal{L}(d|\mu) = \int \mathcal{L}(d|\mu, \theta)\pi(\theta|\mu)d\theta$$

LIGO Noise Properties

- The data consists of both a noise contribution and an astrophysical component

$$\tilde{d}(f) = \tilde{n}(f) + \tilde{h}(\boldsymbol{\theta}; f)$$

noise

astrophysical contribution

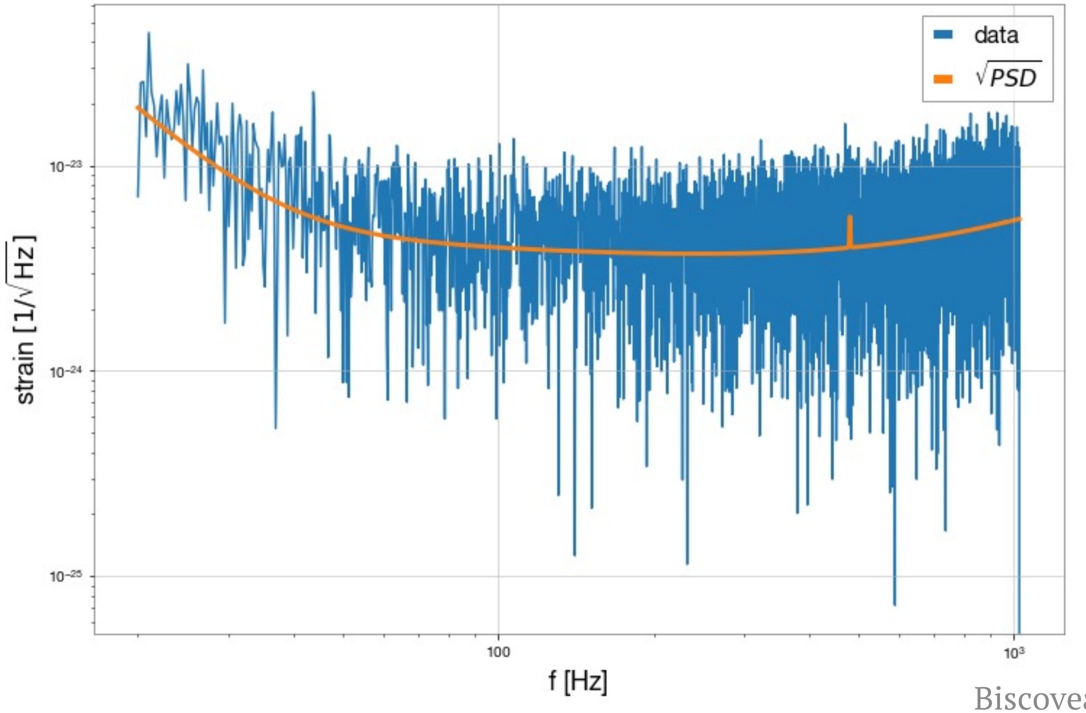
- The noise is typically assumed to be stationary and Gaussian and is characterized by the power spectral density (PSD)

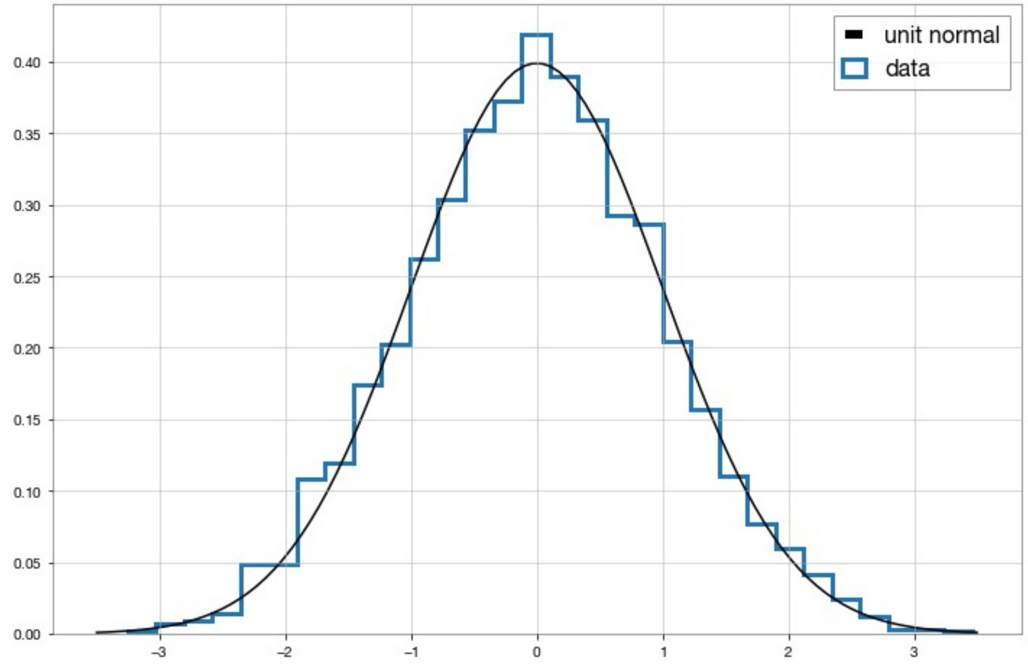
$$\langle \tilde{n}^*(f_i)\tilde{n}(f_j)\rangle = \frac{T}{2} S_n(f)\delta_{ij}$$

- T is the segment duration, PSD has units of [1/Hz]

LIGO Noise Properties

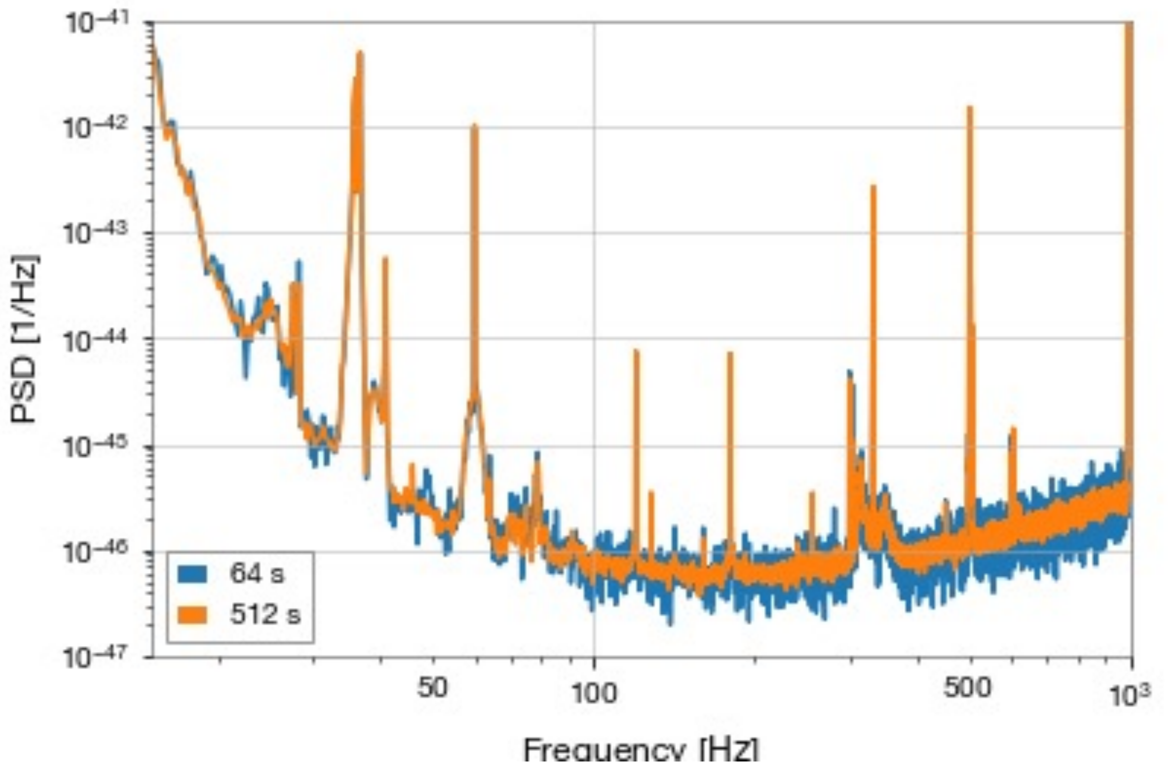
- For well-behaved noise in the absence of a signal, the real and imaginary parts of the strain each follow a unit Gaussian distribution about the square root of the PSD (the amplitude spectral density, ASD)
- Data is whitened by dividing by the ASD





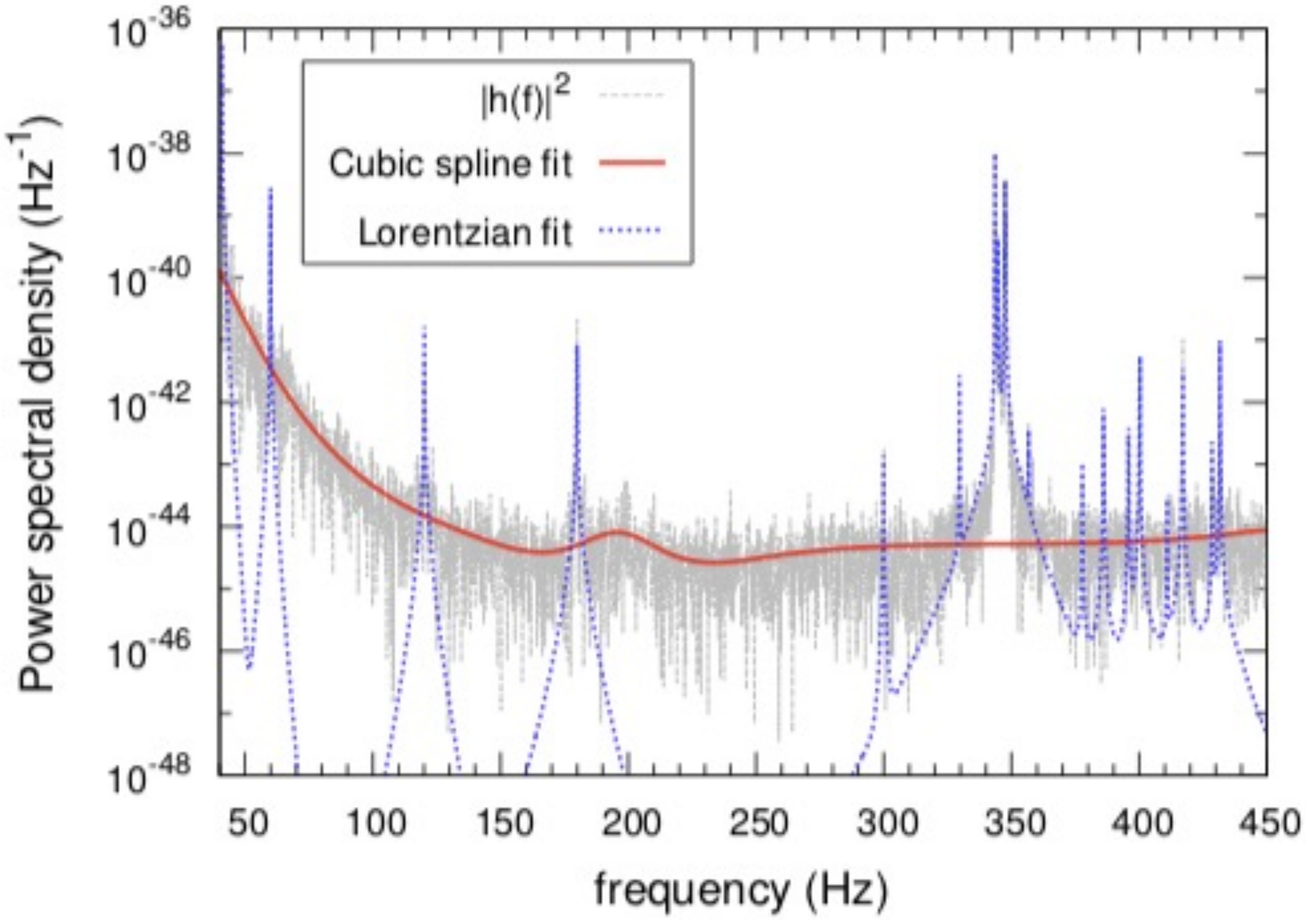
Calculating the PSD

- Off-source method
  - Also called the periodogram method or Welch method
  - Use a long stretch of data either before or after but always excluding the analysis segment
  - Split the data into short segments and calculate $|\tilde{d}(f_i)|^2$ for each segment after windowing the data
  - Take the median of the periodograms from each short data segment



Calculating the PSD

- On-source method
  - Model the PSD as a sum of a broadband spline and narrowband Lorentzians using the BayesLine algorithm
  - Using only the data from the analysis segment, infer the spline and Lorentzian parameters that best characterize the PSD
  - Requires significantly less data → more likely that it will be stationary and Gaussian over a shorter period of time



Littenberg and Cornish (2014)

The Gravitational-Wave Likelihood

- The residual is the difference between the data and the signal template:

$$\tilde{r}(\boldsymbol{\theta}; f_i) = \tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)$$

- In the presence of a signal, the real and imaginary parts of the residual should also be Gaussian-distributed:

$$p(\Re\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\Re\tilde{r}(\boldsymbol{\theta}; f_i))^2}{2\sigma_i^2}\right)$$

$$p(\Im\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\Im\tilde{r}(\boldsymbol{\theta}; f_i))^2}{2\sigma_i^2}\right)$$

The Gravitational-Wave Likelihood

- The PSD (and some normalization factors) is the variance of the likelihood:

$$\sigma_i^2 = \frac{T S_n(f_i)}{4}$$

- The total likelihood of the data is the product of the real and imaginary likelihoods:

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}) \equiv p(\Re\tilde{d}(f_i)|\boldsymbol{\theta})p(\Im\tilde{d}(f_i)|\boldsymbol{\theta})$$

The Whittle Likelihood

- The final form of the likelihood used in gravitational-wave data analysis is the Whittle Likelihood:

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{2}{T\pi S_n(f_i)} \exp\left(-\frac{2|\tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)|^2}{T\ S_n(f_i)}\right)$$

$$\mathcal{L}(d|\boldsymbol{\theta}) = \prod_i \mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta})$$

- The likelihood for multiple frequencies is the product of the individual-frequency likelihoods, same for multiple detectors:

$$\mathcal{L}(\{d\}_j|\boldsymbol{\theta}) = \prod_j \mathcal{L}(d_j|\boldsymbol{\theta})$$

Noise-weighted inner product

- Define the noise-weighted inner product:

$$\langle a|b\rangle \equiv \frac{4}{T}\sum_{j}\Re\left(\frac{\tilde{a}(f_i)^*\tilde{b}(f_i)}{S_n(f_i)}\right)$$

- This will be useful for simplifying the likelihood and defining various signal-to-noise ratios

Rearranging the log-likelihood

$$\log \mathcal{L}(d|\boldsymbol{\theta}) = \sum_i \log \mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta})$$

$$\log \mathcal{L}(d|\boldsymbol{\theta}) = \sum_i \log \left(\frac{2}{T\pi S_n(f_i)} \right) - \sum_i \frac{2|\tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)|^2}{T S_n(f_i)}$$

$$\log \mathcal{L}(d|\boldsymbol{\theta}) \propto -\frac{1}{2} \langle d - h(\boldsymbol{\theta}) | d - h(\boldsymbol{\theta}) \rangle$$

- Log-likelihood is proportional to the noise-weighted inner product of the residual with itself

Signal-to-noise ratios

- We can break the likelihood down into some special terms:

$$\log \mathcal{L}(d|\boldsymbol{\theta}) \propto -\frac{1}{2}\left[\langle d|d\rangle - 2\langle d|h(\boldsymbol{\theta})\rangle + \langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle\right]$$

- Optimal signal-to-noise ratio:

$$\rho^2_{\text{opt}} \equiv \langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle$$

- Matched filter signal-to-noise ratio:

$$\rho_{\text{MF}} = \frac{\langle d|h(\boldsymbol{\theta})\rangle}{\rho_{\text{opt}}}$$

Noise likelihood

- There is a hypothesis implicit in the likelihood presented so far—that there is a signal in the data
- In the absence of a signal, the likelihood simplifies

$$\mathcal{L}(\tilde{d}(f_i)|\emptyset) = \frac{2}{T\pi S_n(f_i)} \exp\left(-\frac{2|\tilde{d}(f_i)|^2}{T\ S_n(f_i)}\right)$$

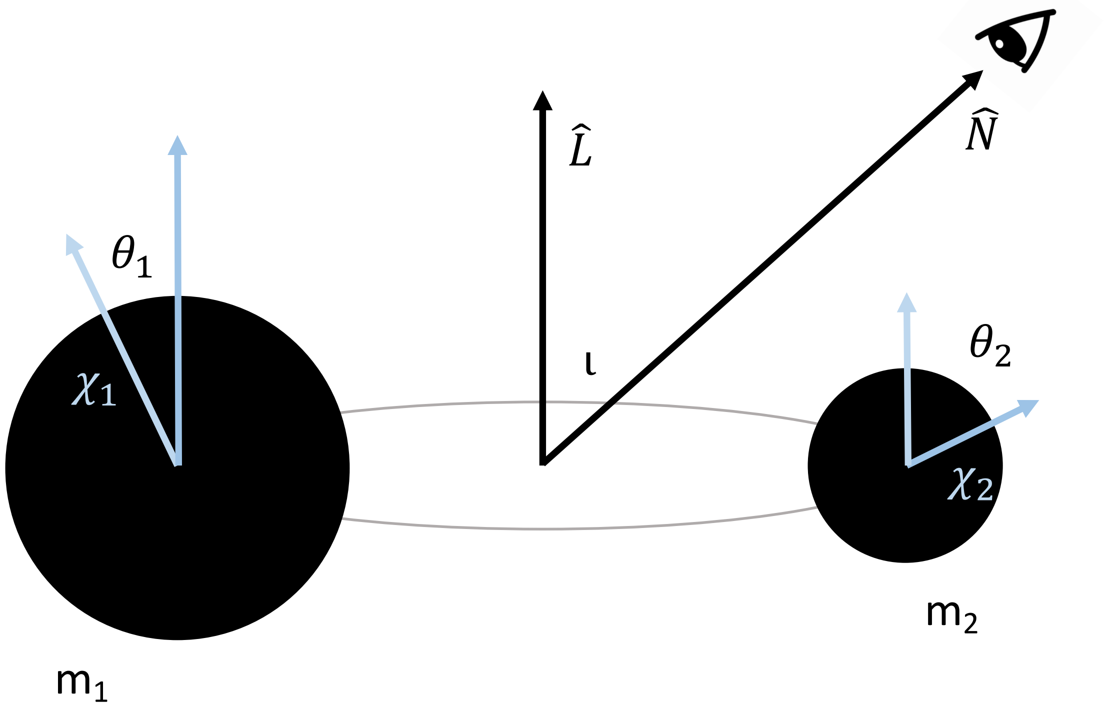
$$\log \mathcal{L}(d|\emptyset) \propto -\frac{1}{2}\langle d|d\rangle$$

- This is called the noise likelihood

The astrophysical contribution

- For compact binary coalescences, the astrophysical contribution, $\tilde{h}(\boldsymbol{\theta}: f_i)$, is a waveform that depends on 17 parameters

Intrinsic:
Component masses
Component spins
(Tidal deformabilities)



Extrinsic:
Sky location
Distance
Inclination
Polarization
Reference phase
Time at coalescence

- Other models for other types of signals – sine gaussian wavelets, supernova waveforms, etc.

Measuring source properties from the waveform

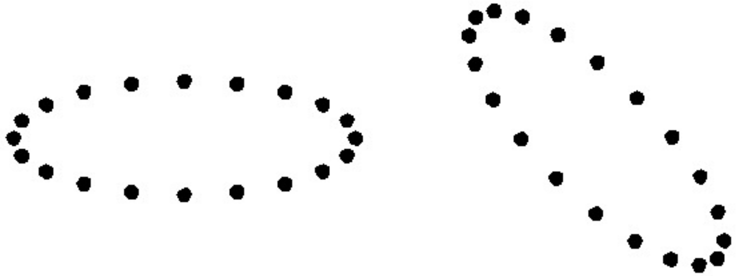
$$\boldsymbol{\theta} \in [m_1, m_2, \vec{\chi}_1, \vec{\chi}_2, (\Lambda_1, \Lambda_2), \iota, d_L, \alpha, \delta, \psi, \phi, t_c]$$

- The intrinsic parameters and distance affect the amplitude and phase of the waveform

$$\tilde{h}_+(f) = \frac{1}{2}\mathcal{A}_{\rm GW}(f)(1+\cos^2\iota)\cos\phi_{\rm GW}(f)$$

$$\tilde{h}_\times(f) = \mathcal{A}_{\rm GW}(f)\cos\iota\sin\phi_{\rm GW}(f)$$

- Two polarizations – plus and cross

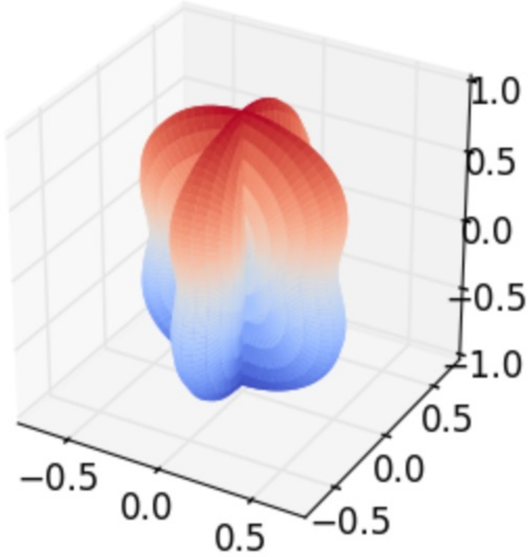


https://www.ligo-india.in/wp-content/uploads/2021/01/3-Stretch-and-Squeeze2.gif

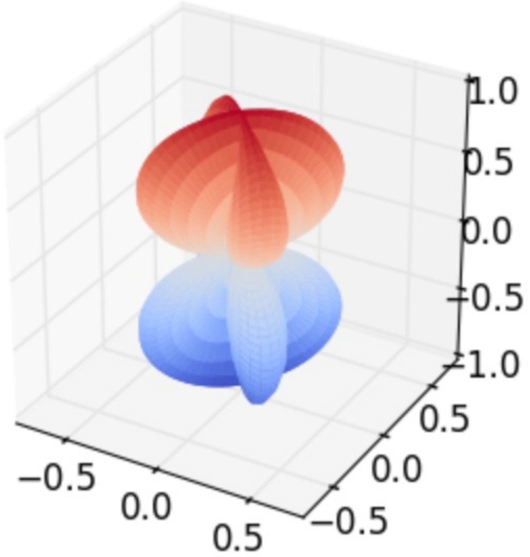
Detector response

- The detector's sensitivity to the plus and cross modes of the gravitational wave depends on its orientation and geometry via the antenna pattern functions

$$\tilde{h}(f) = F_+(\alpha, \delta, \psi)\tilde{h}_+(f) + F_\times(\alpha, \delta, \psi)\tilde{h}_\times(f)$$



Plus



Cross

Mass parameters

- "Primary" refers to more massive black hole, "secondary" to least massive
- Best-measured mass parameter with gravitational waves is the chirp mass:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

- Symmetric mass ratio: $\eta = \frac{m_1 m_2}{(m_1 + m_2)^2}$
- Asymmetric mass ratio: $q = m_2/m_1$

Mass parameters

- Masses indicative of supernova physics:
  - Some theoretical models (and observations of galactic neutron stars and black holes) predict a lower mass gap between the most massive neutron stars and least massive black holes formed in supernovae
  - Upper mass gap predicted since black holes formed by gravitational collapse are not expected to be more massive than ~50 solar masses due to (pulsational) pair-instability supernovae

GW190814

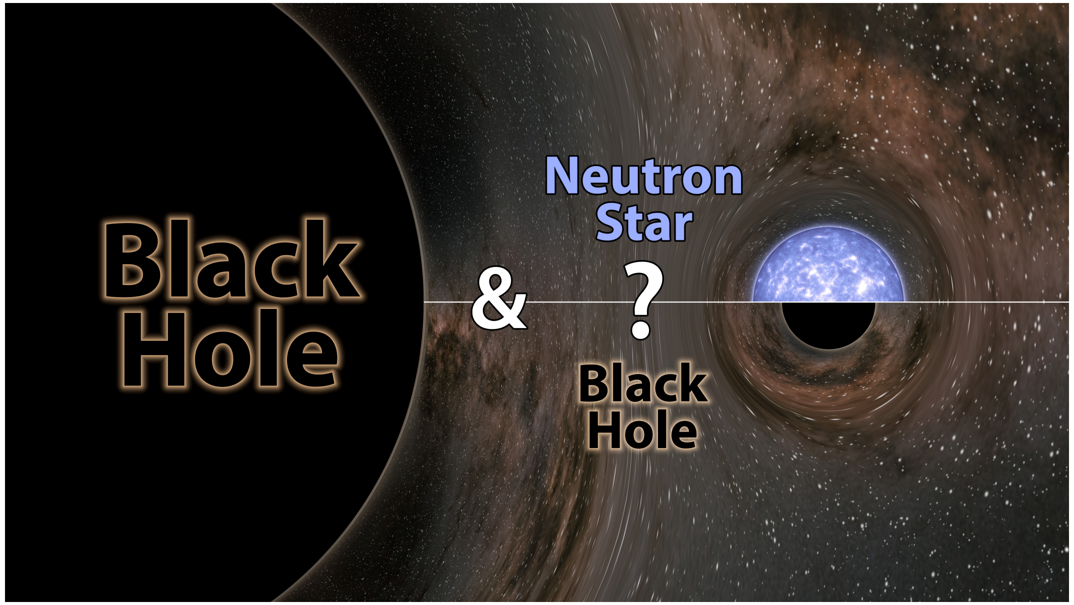
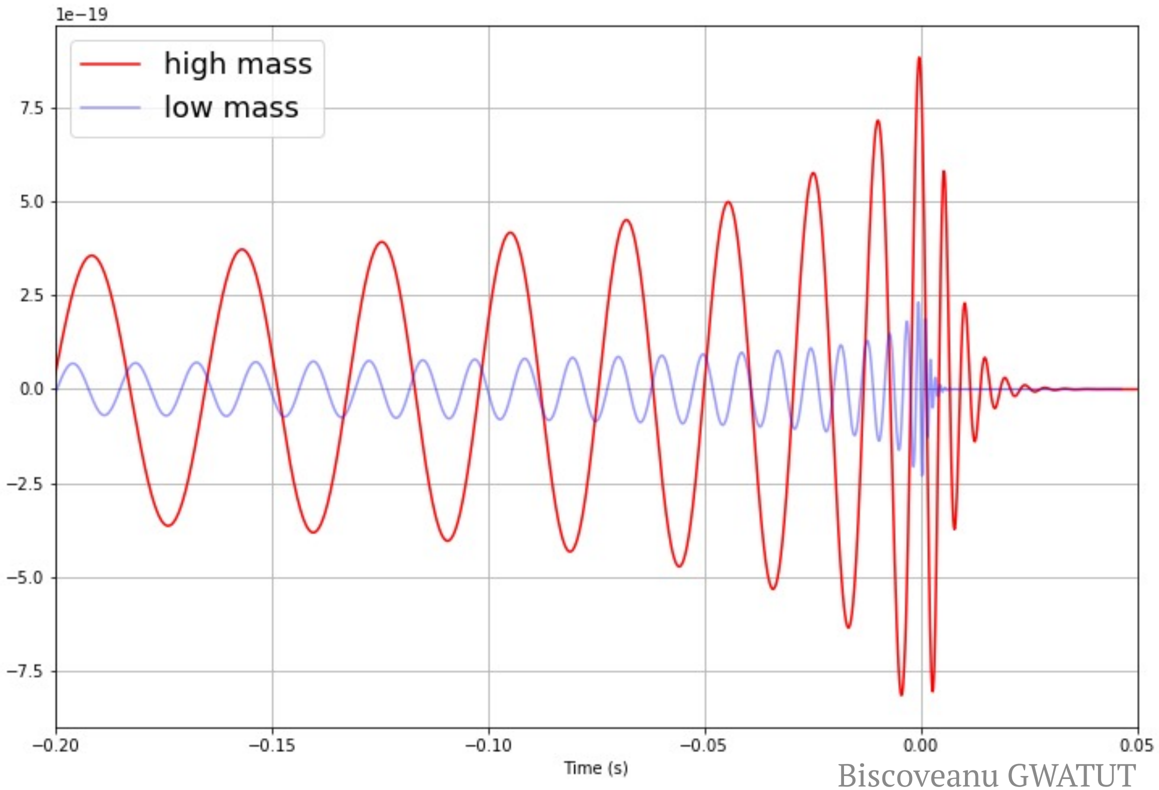


Image credit: LIGO/Caltech/MIT/R. Hurt (IPAC).

Effect of Mass

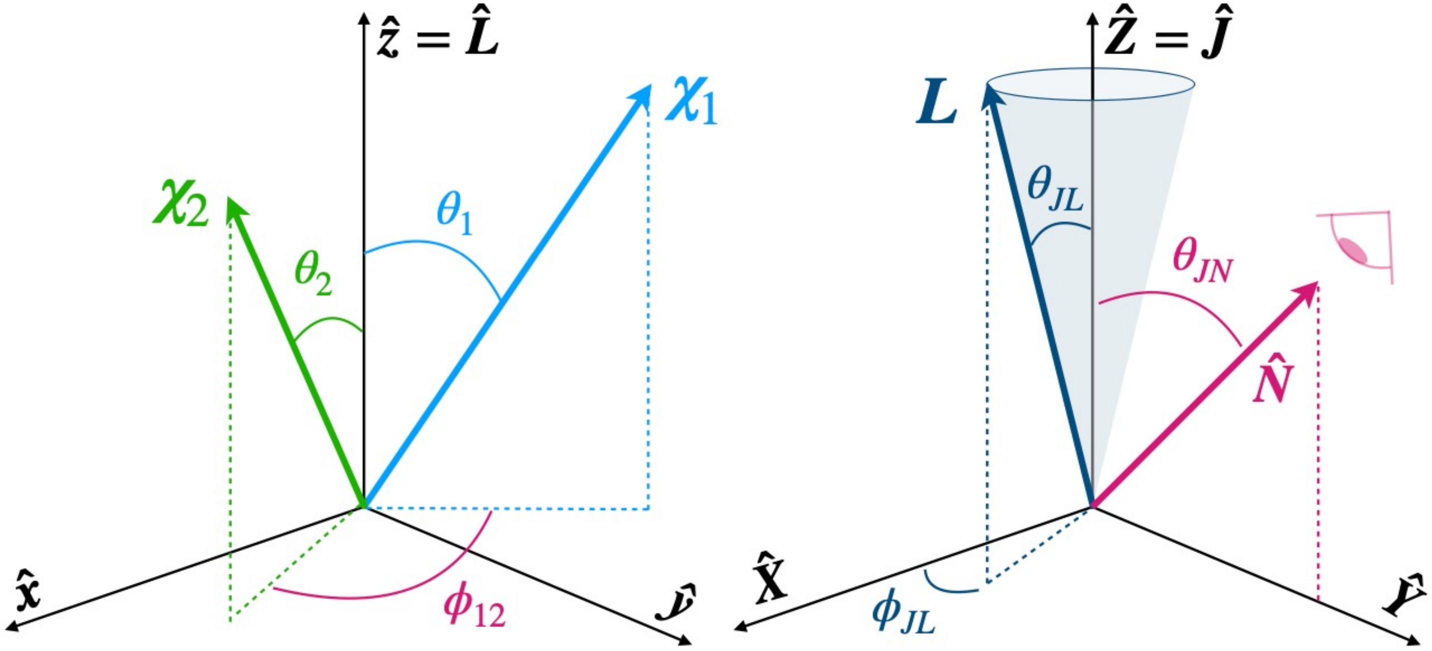
- Bigger mass → bigger amplitude
- Final mass measured from ringdown



$$\mathcal{A}_{\rm GW} \propto \frac{\mathcal{M}^{5/6} f^{-7/6}}{d_L}$$

Spin Parameters

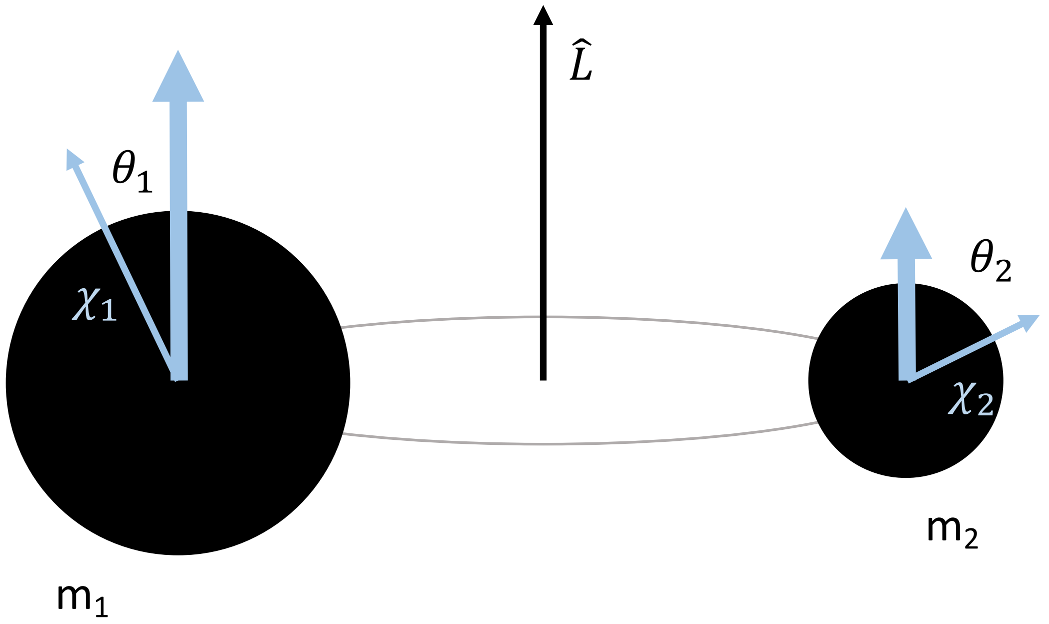
- Misalignment between the spins and the orbital angular momentum causes precession
  - Total angular momentum - $\hat{J}$
  - Orbital angular momentum - $\hat{L}$
  - Observer line of sight - $\hat{N}$



Spin Parameters

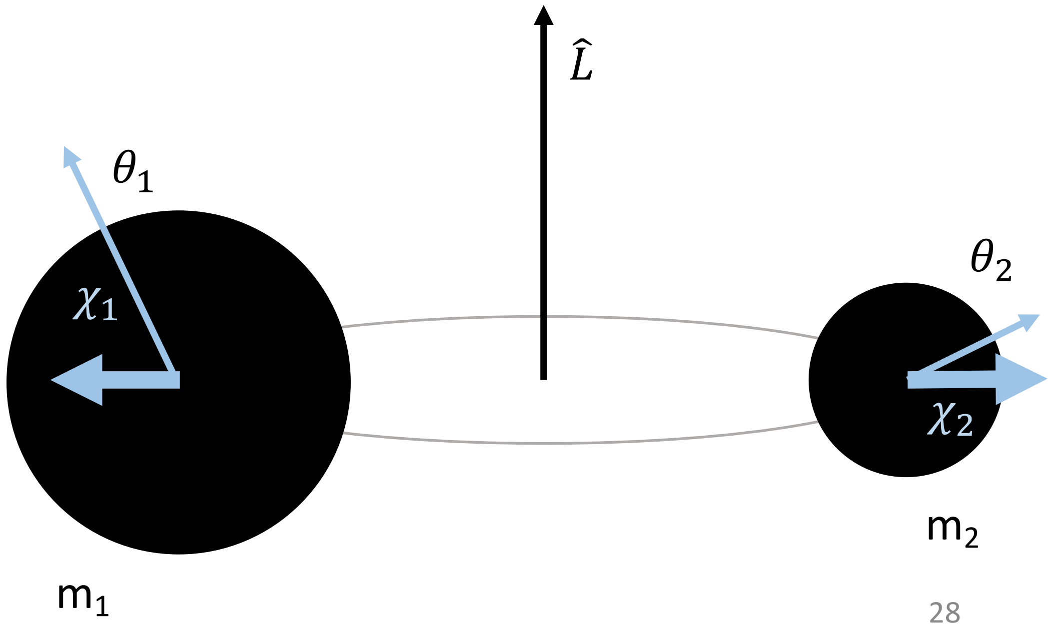
- χ_{eff} - best measured spin parameter with gravitational waves, mass-weighted spin aligned with the orbital angular momentum

$$\chi_{\text{eff}} = \frac{\chi_1 \cos\theta_1 + q\chi_2 \cos\theta_2}{1+q}$$



- χ_p - effective precessing spin, mass-weighted spin projection onto the orbital plane

$$\chi_p = \max\left(\chi_1 \sin\theta_1, \left(\frac{4q+3}{4+3q}\right) q\chi_2 \sin\theta_2\right)$$



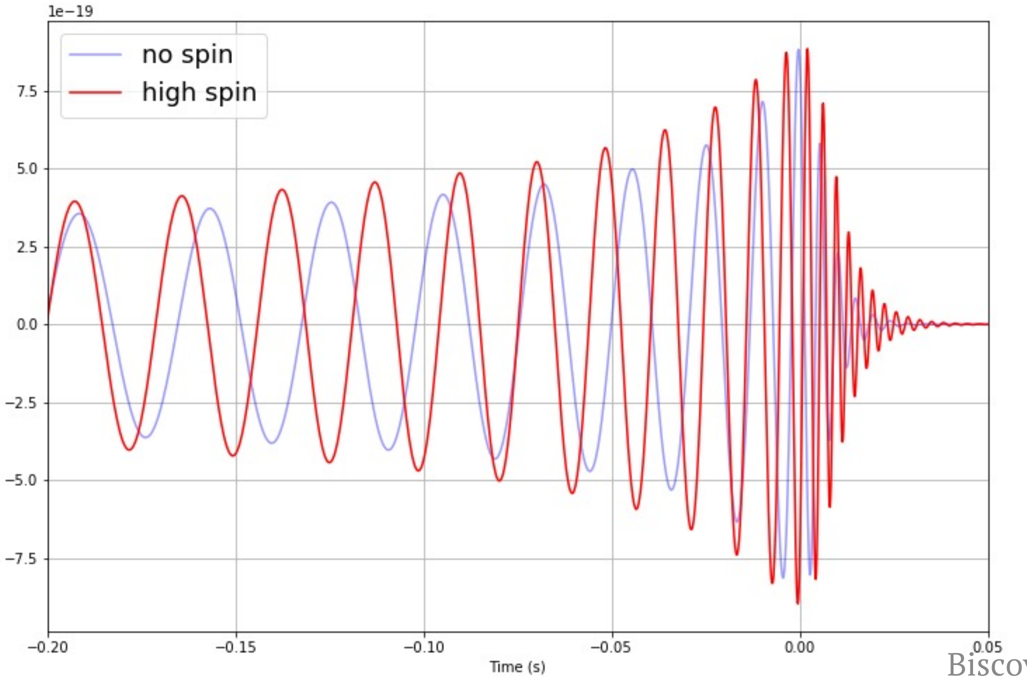
Spin parameters

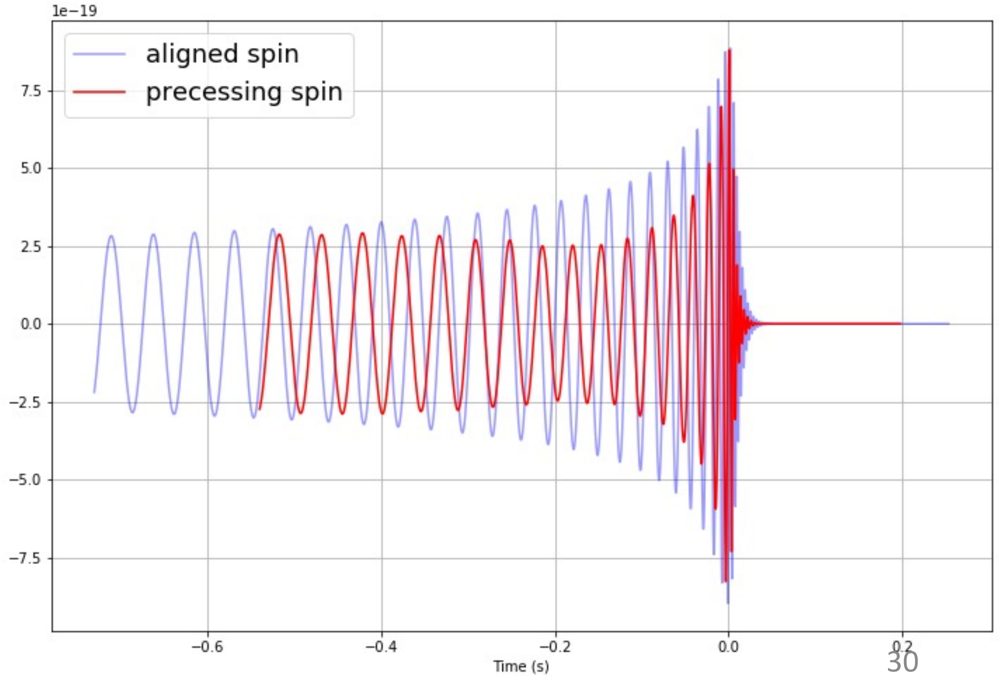
- Tilts are an indicator of binary formation channel:
  - Binary stars evolving to compact objects in isolation expected to have aligned spins
  - Compact binaries assembled dynamically in dense environments expected to have isotropic spin distribution
- Azimuthal angle ϕ_{12} can be indicative of efficiency of tidal interactions during stellar evolution via spin-orbit resonances
- Both azimuthal angles critical for determining the kick of the remnant object→ retention fraction

Effect of Spin

- More positive aligned spin → orbital hangup
- Takes longer for the system to merge

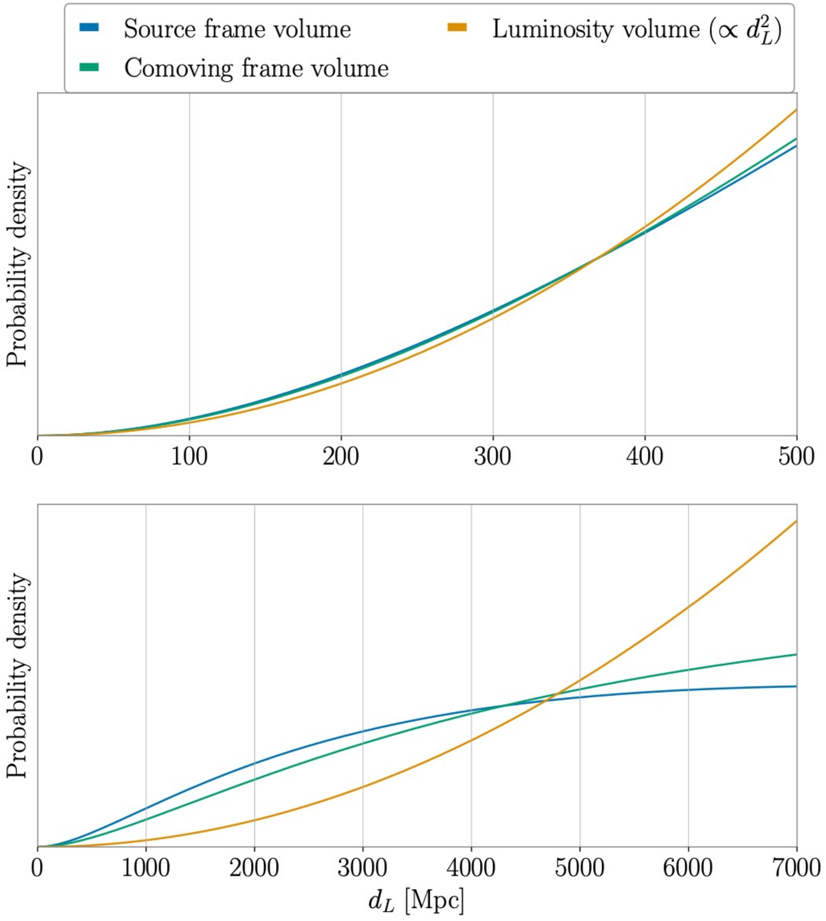
- Precessing spins → amplitude modulations
- Spins misaligned to orbital angular momentum





Priors

- Uniform in some parameterization of the mass
  - Easier to sample in chirp mass and mass ratio
- Enforce $m_1 > m_2$
- Uniform in spin magnitudes
- Spin angles isotropic on the sphere
- Isotropic on the sky for right ascension and declination
- Uniform in luminosity volume ($\propto d_L^2$)
  - Or cosmological distance prior



Waveforms

- See Patricia Schmidt's talk tomorrow!
- Different methods for providing approximate solutions for $\tilde{h}(\boldsymbol{\theta}; f)$:
  - Post-Newtonian, inspiral-only
  - Inspiral-merger-ringdown phenomenological waveforms tuned to numerical relativity
  - Effective one-body or self-force methods
  - Numerical relativity surrogate models
- Different models include different physics:
  - Tides, precession, higher-order modes

Bayesian Model Selection

- Simple example – signal versus noise
- Noise evidence: integrate the noise likelihood (no astrophysical contribution) over the binary parameters:

$$\begin{aligned}\mathcal{Z}_N &= \int \mathcal{L}(d|\emptyset)\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \mathcal{L}(d|\emptyset)\int \pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \mathcal{L}(d|\emptyset)\end{aligned}$$

Bayesian Model Selection

- Bayes factor: evidence ratio

$$\mathrm{BF}_N^S = \frac{\mathcal{Z}_S}{\mathcal{Z}_N}$$

- Odds ratio: bayes factor weighted by prior odds

$$\mathcal{O}_N^S = \mathrm{BF}_N^S \frac{\pi(S)}{\pi(N)}$$

Bayesian Model Selection

- Another example – aligned vs precessing spins

$$\mathrm{BF}^{A}_{P} = \frac{\int \mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|A)\ d\theta}{\int \mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|P)\ d\theta}$$

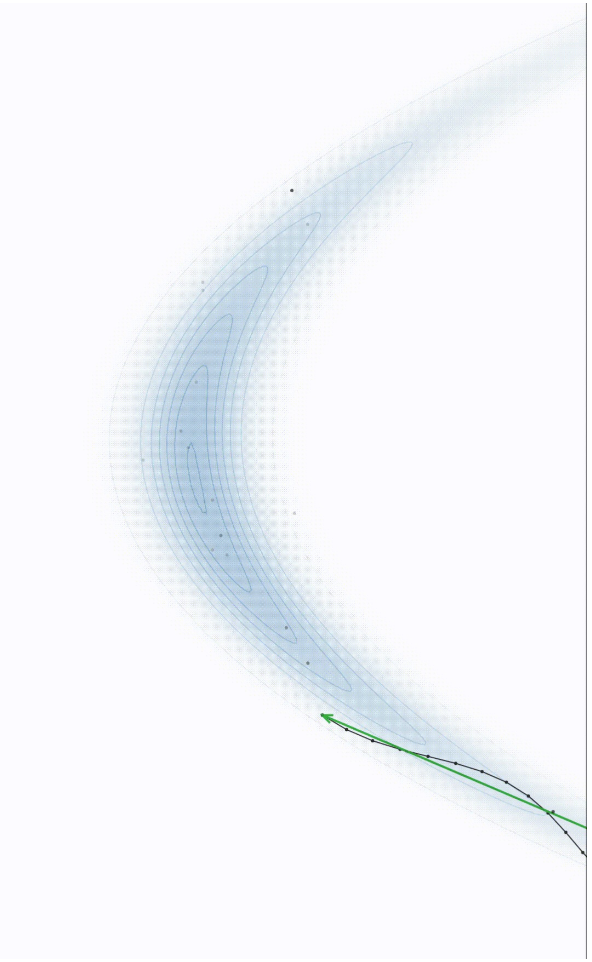
- For aligned spins, prior is a delta function at zero on tilt angles
- Typically BF > 3000 is significant

Sampling methods

- How do you actually obtain $p(\theta|d)$?
- Could evaluate the likelihood on a grid, but this isn't feasible with 17 parameters
- Instead use a stochastic sampler:
  - Markov Chain Monte Carlo (MCMC)
  - Nested sampling
- Obtain samples from the posterior probability distribution

MCMC

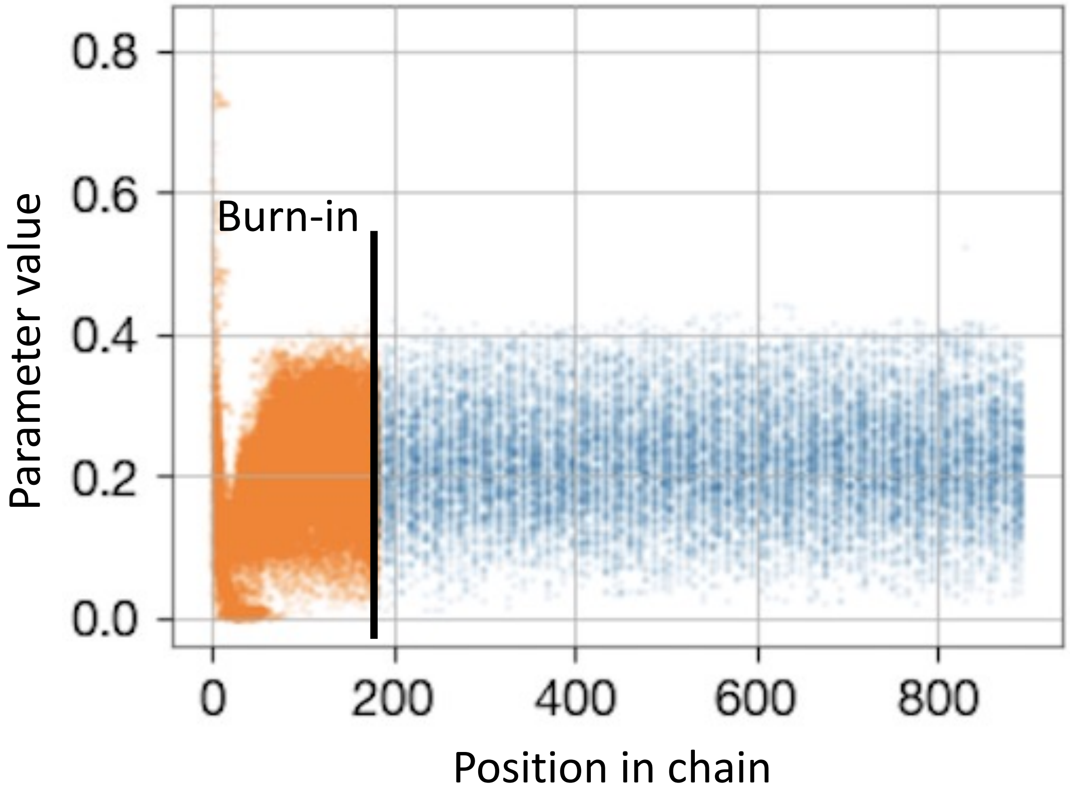
- Particles undergo a random walk through the parameter space, where the probability of jumping to a new location is dictated by the proposal density function
- Proposed sample is accepted with probability depending on ratio of the product of likelihood, prior, and proposal distribution at the old vs new points
- Determining a suitable proposal density function is the hard part of sampling – a simple example is a Gaussian centered on the current location
- Proposal should maintain detailed balance—probability of jumping between two points in parameter space is the same in both directions



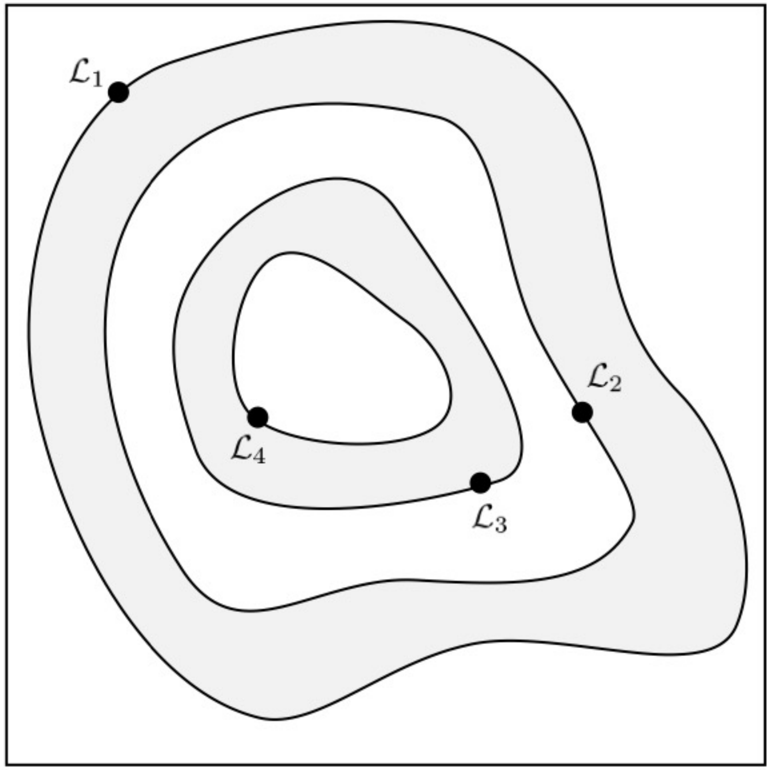
https://github.com/chi-feng/mcmc-demo

MCMC

- Burn-in period before the walkers "forget" their starting positions
- Adjacent samples in a chain are correlated – chains need to be thinned by the integrated autocorrelation time (ACT)
- Number of steps the walker must take before it "forgets" where it came from
- Smaller ACT means sampler converges faster, depends on efficiency of proposal distribution



Nested Sampling



(a)

F. Feroz et. al. (2008)

- Sprinkle a set of live points over the prior space
- Replace the live point with the lowest likelihood with a point with a higher likelihood
- Evidence is the product of the likelihood at the discarded point and the difference in the prior volume between iterations
- Obtain samples from the prior in the process of calculating the evidence
- Proceed until a termination criterion is reached

Nested Sampling

- Evidence is the expectation value of the likelihood, which can be obtained by integrating over (1 - the CDF of the likelihood):

$$
\begin{aligned}
\mathcal{Z} &= \int \mathcal{L}(\theta)\pi(\theta)d\theta \\
&= \langle \mathcal{L}(\theta) \rangle \\
&= \int_0^\infty (1 - \underbrace{F_{\mathcal{L}}}_{\text{CDF}}(\underbrace{L}_{\text{Some value of the likelihood}}))dL
\end{aligned}
$$

Nested Sampling

- CDF of the likelihood, F(L), is the area enclosed in regions of parameter space where likelihood < L:

$$F_{\mathcal{L}}(L) = \int_{\mathcal{L}<L} \pi(\theta) d\theta$$

- Define the prior volume:

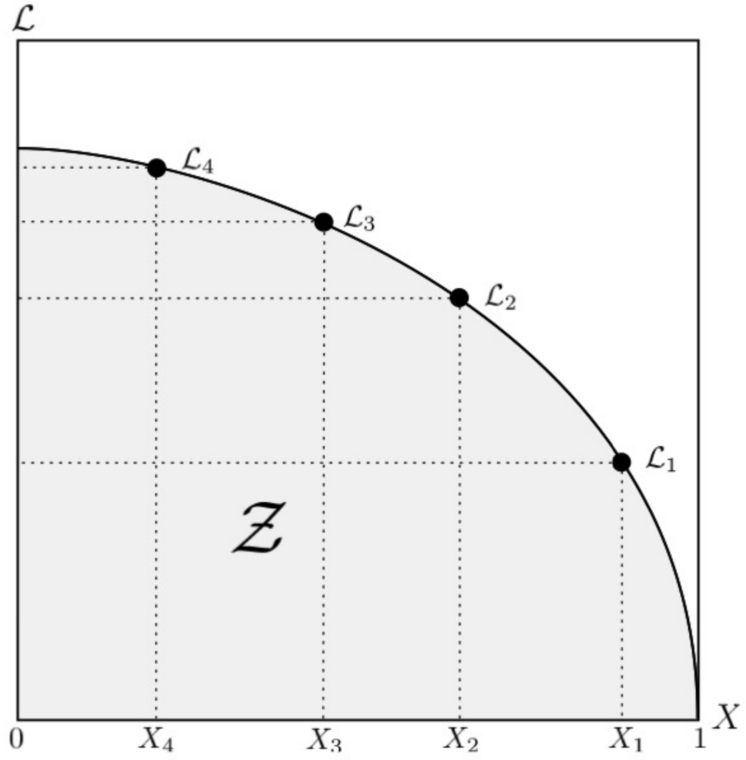
$$X(L) \equiv 1 - F_{\mathcal{L}}(L) = \int_{\mathcal{L}>L} \pi(\theta) d\theta$$

- X(L) is a monotonically increasing function between X(0) = 1 (integral over full normalized prior) to $X(L_{max})=0$

Nested Sampling

- We can now turn the evidence into a one-dimensional integral:

$$\mathcal{Z} = \int_0^\infty X(L)dL$$
$$= \int_0^1 L(X)dX$$



(b)

F. Feroz et. al. (2008)

- For the i^{th} iteration:

$$L_{\min} = L\left(\left(\frac{N-1}{N}\right)^i\right)$$

Bilby

- The **B**ayesian **I**nference **Lib**ra**ry** is a software package designed to enable parameter estimation for compact binary coalescences and more general problems
- Emphasis on modularity, transparency, and ease of use
- Wrapper for many different external samplers including dynesty, pymultinest, cpnest, emcee, ptemcee, and others
- Can analyze real data from LIGO and Virgo or simulated signals

Additional Resources

- https://lscsoft.docs.ligo.org/bilby/ - Bilby documentation
- https://chi-feng.github.io/mcmc-demo/ - cool animations of MCMC
- Further reading:
  - Veitch et. al. (2015) https://arxiv.org/pdf/1409.7215.pdf
  - Ashton et. al. (2018) https://arxiv.org/abs/1811.02042
  - Thrane and Talbot (2019) https://arxiv.org/pdf/1809.02293.pdf

Off-Source Method

- Two corrections need to be applied:
  - Window factor to correct for power lost to window, w_t:

$$W = \frac{1}{N_t} \sum_{t=0}^{N_t - 1} w_t^2$$

  - Median correction, where ℓ is the segment number

$$\alpha = \sum_{\ell=1}^{N_s} \frac{(-1)^{\ell}}{\ell}$$