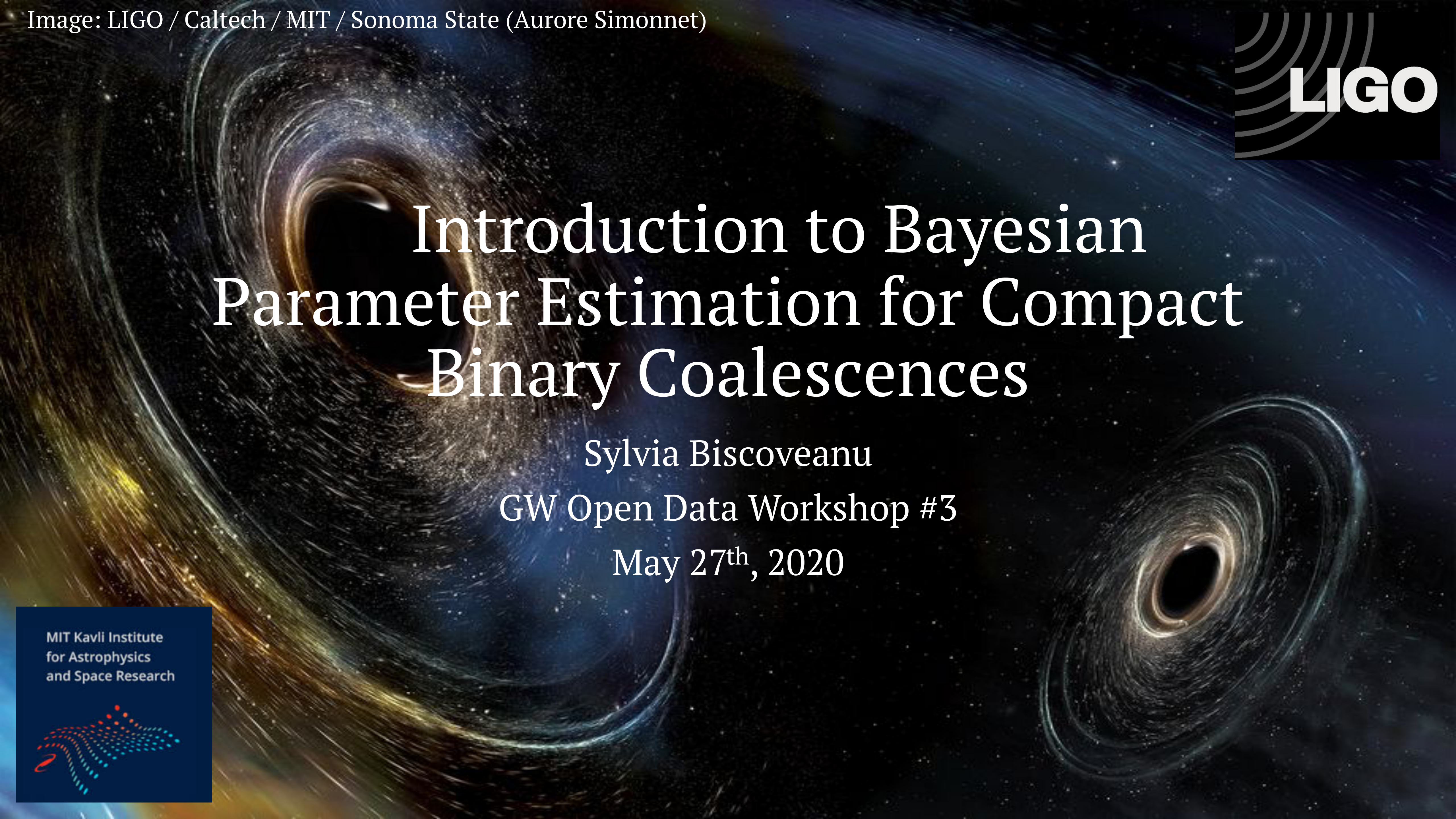
Image: LIGO / Caltech / MIT / Sonoma State (Aurore Simonnet)



Introduction to Bayesian Parameter Estimation for Compact Binary Coalescences

Sylvia Biscoveanu

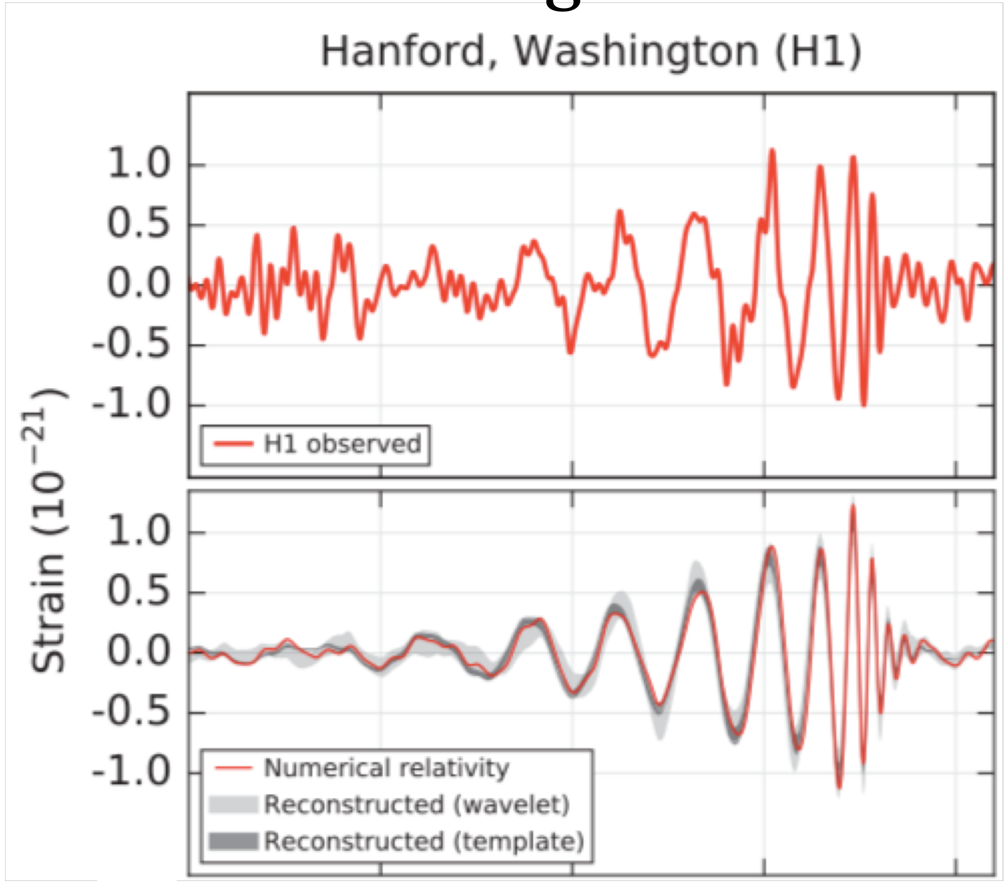
GW Open Data Workshop #3

May 27th, 2020

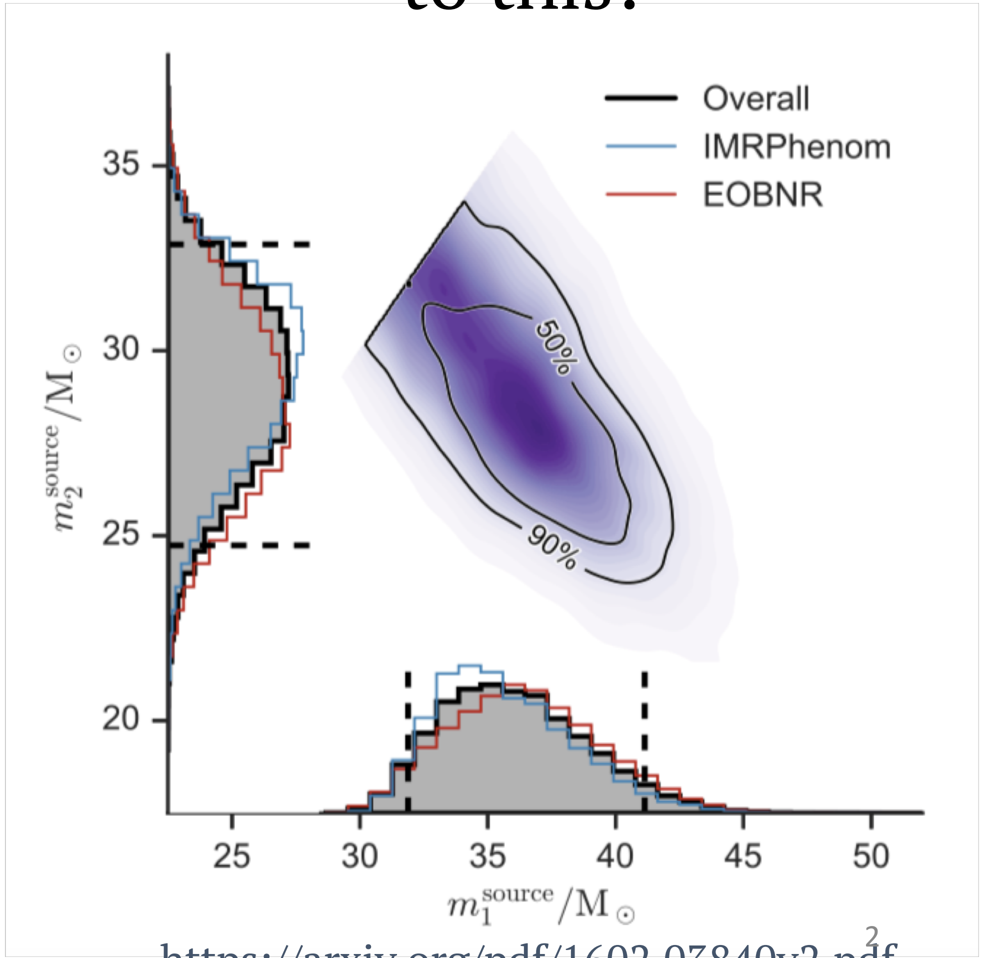
MIT Kavli Institute
for Astrophysics
and Space Research

Motivation

- How do we go from this… to this?



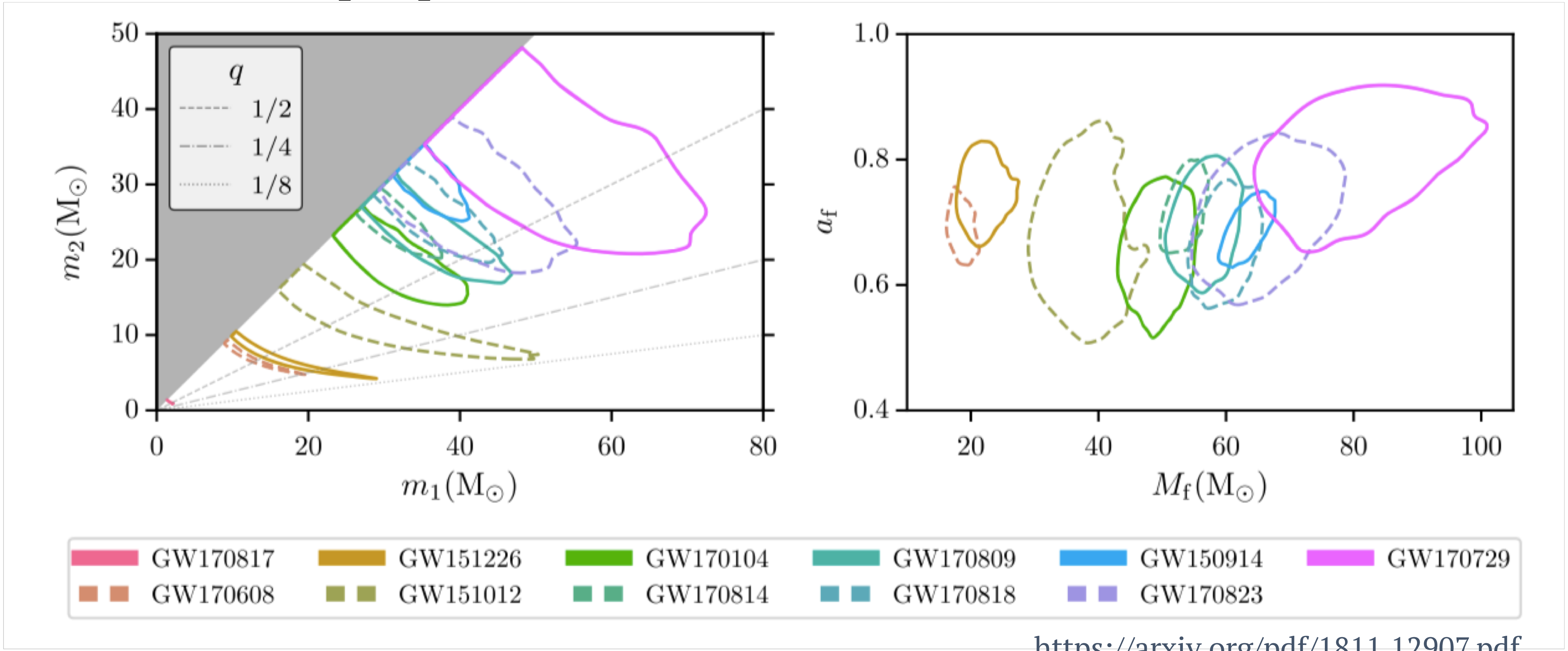
https://arxiv.org/abs/1602.03837



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

Motivation

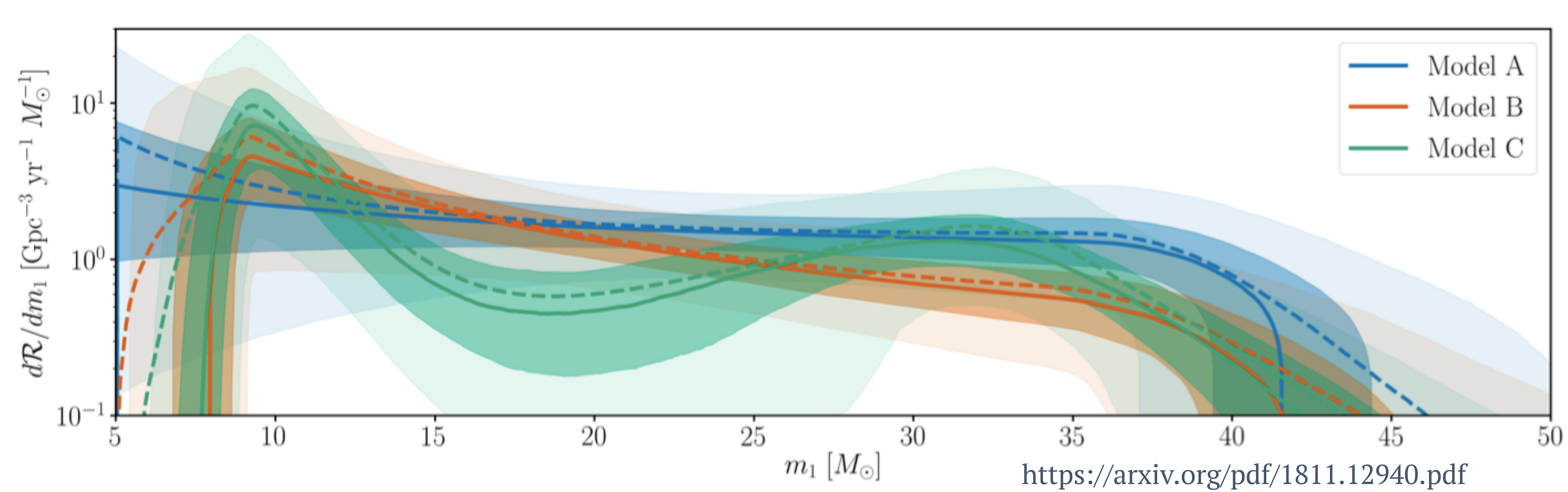
- Measure the properties of individual detections…



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

Motivation

- And of the underlying population



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

Bayes' Theorem

$$\underbrace{p(\boldsymbol{\theta}|d, H)}_{\text{posterior}} = \frac{\overbrace{p(d|\boldsymbol{\theta}, H)}^{\text{likelihood}}\,\overbrace{p(\boldsymbol{\theta}|H)}^{\text{prior}}}{\underbrace{p(d|H)}_{\text{evidence}}}$$

posterior

likelihood

prior

evidence

parameters ($\boldsymbol{\theta}$) data (d) model (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

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

Bayes' Theorem Components

- Evidence – normalization constant for the posterior, marginalized likelihood

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

Putting it all together:

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

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

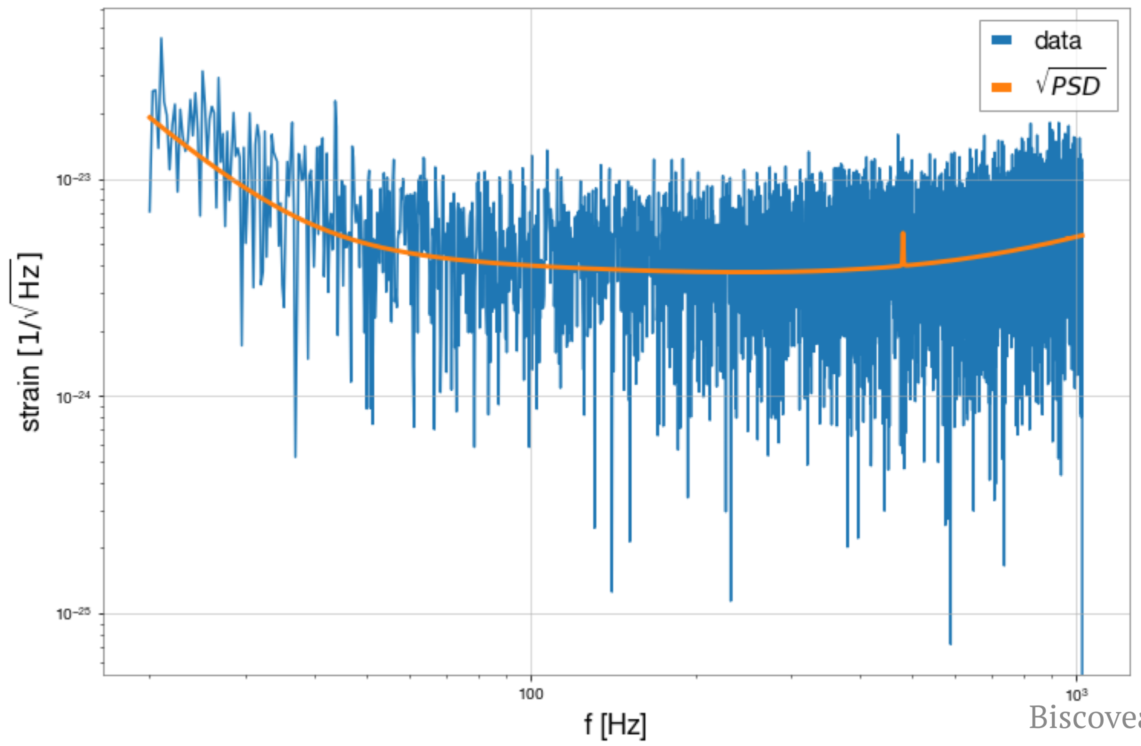
- The noise is typically assumed to be stationary and Gaussian and is characterized by the power spectral density (PSD)

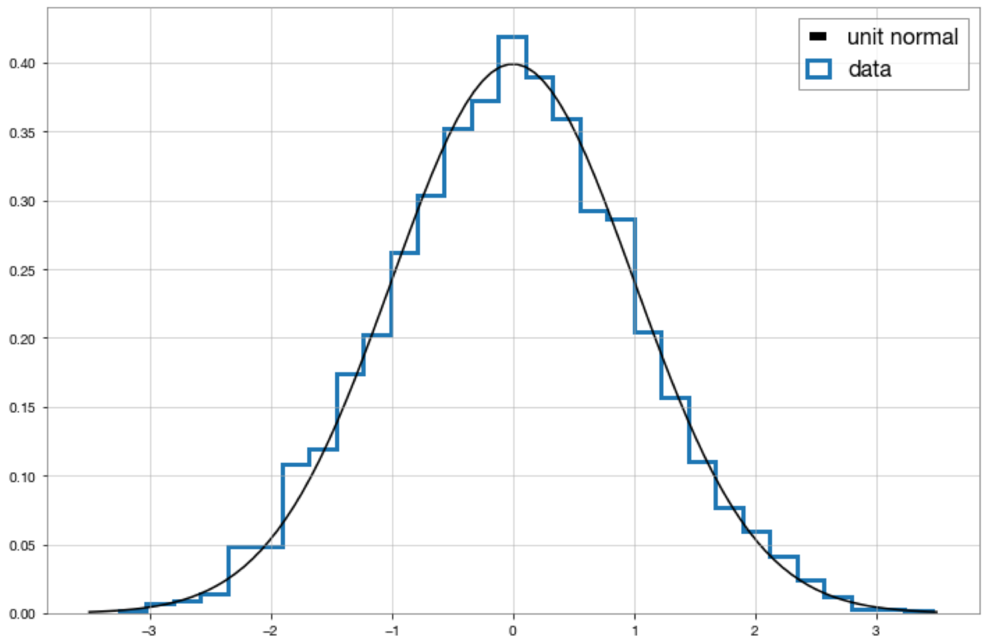
$$\langle \tilde{n}^*(f_i)\tilde{n}(f_j)\rangle = \frac{T}{2} S_n(f)\delta_{ij}$$

- T is the segment duration

LIGO Noise Properties

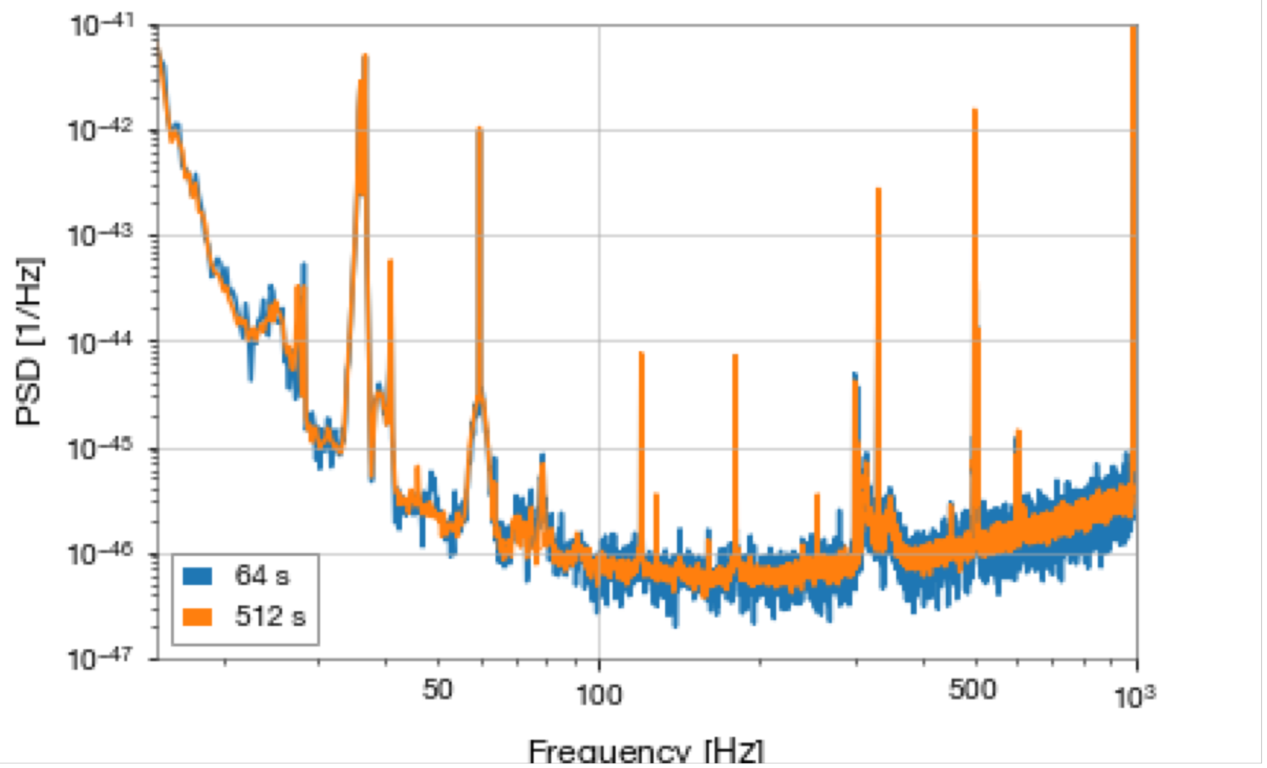
- When the noise is well-behaved, the strain follows a unit Gaussian distribution about the square root of the PSD (in the absence of a signal)





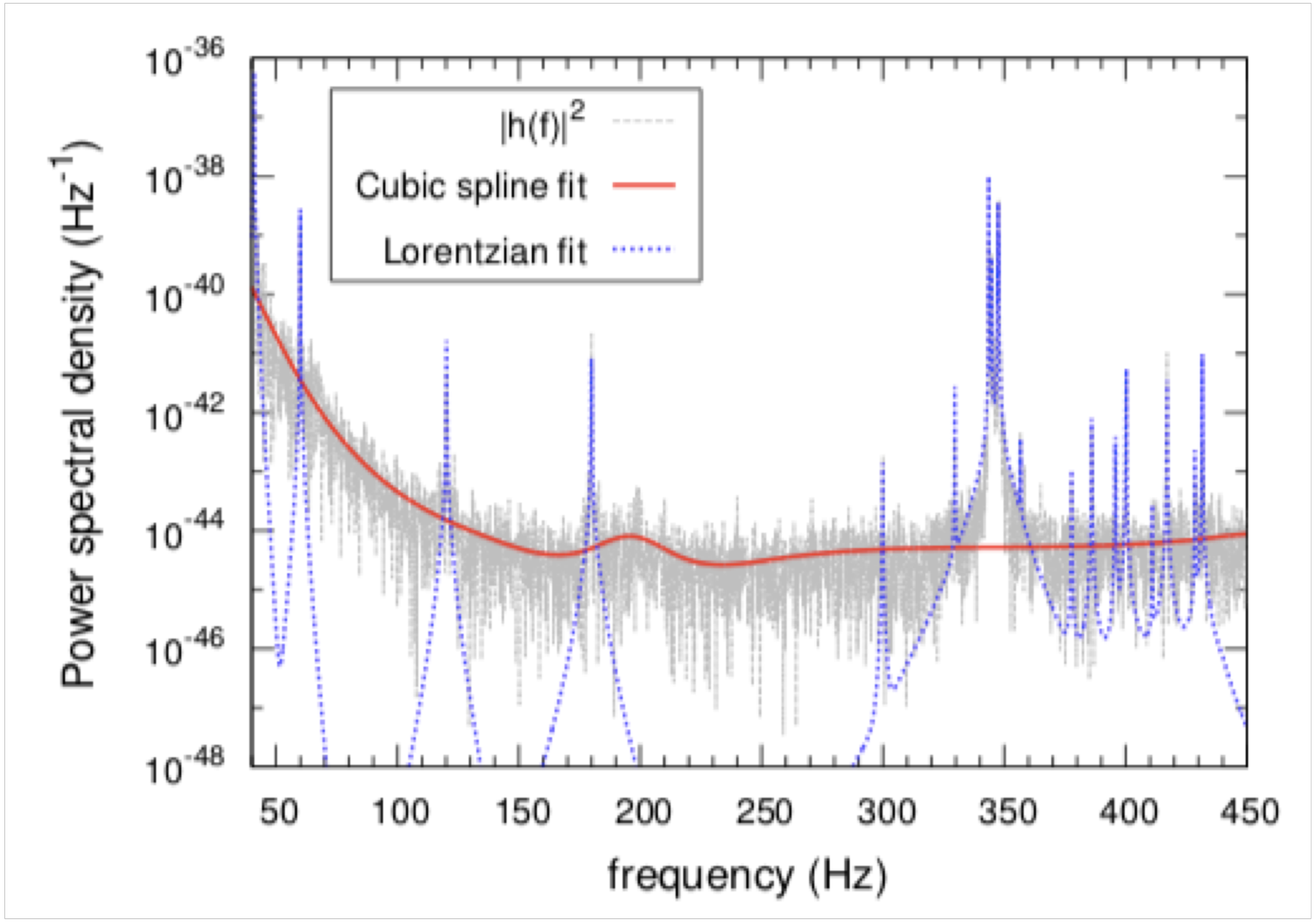
Calculating the PSD

- 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

- In the presence of a signal, the residual between the data and the best-matching template should also follow a unit Gaussian about the square root of the PSD:

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

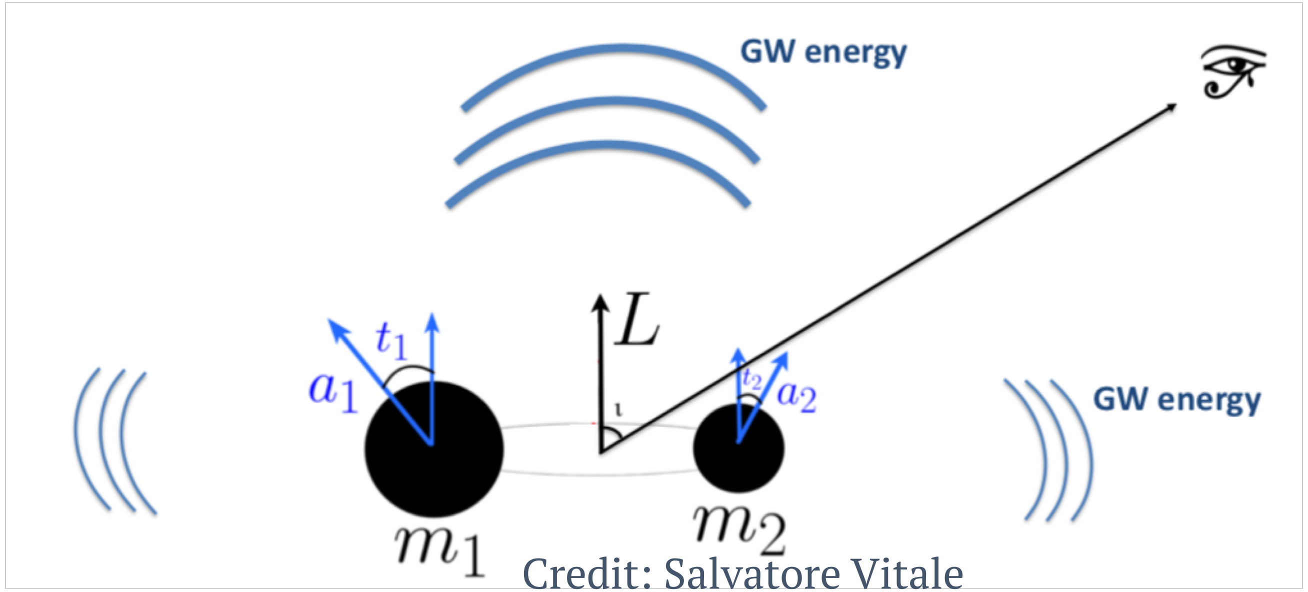
$$\mathcal{L}(d|\boldsymbol{\theta}) = \prod_i \mathcal{L}(d(f_i)|\boldsymbol{\theta})$$

- Total likelihood is the product of the individual frequency bins

The astrophysical contribution

- For compact binary coalescences, the astrophysical contribution is a waveform that depends on 17 parameters

Intrinsic:
Component masses
Component spins
(Tidal deformabilities)



Credit: Salvatore Vitale

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

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

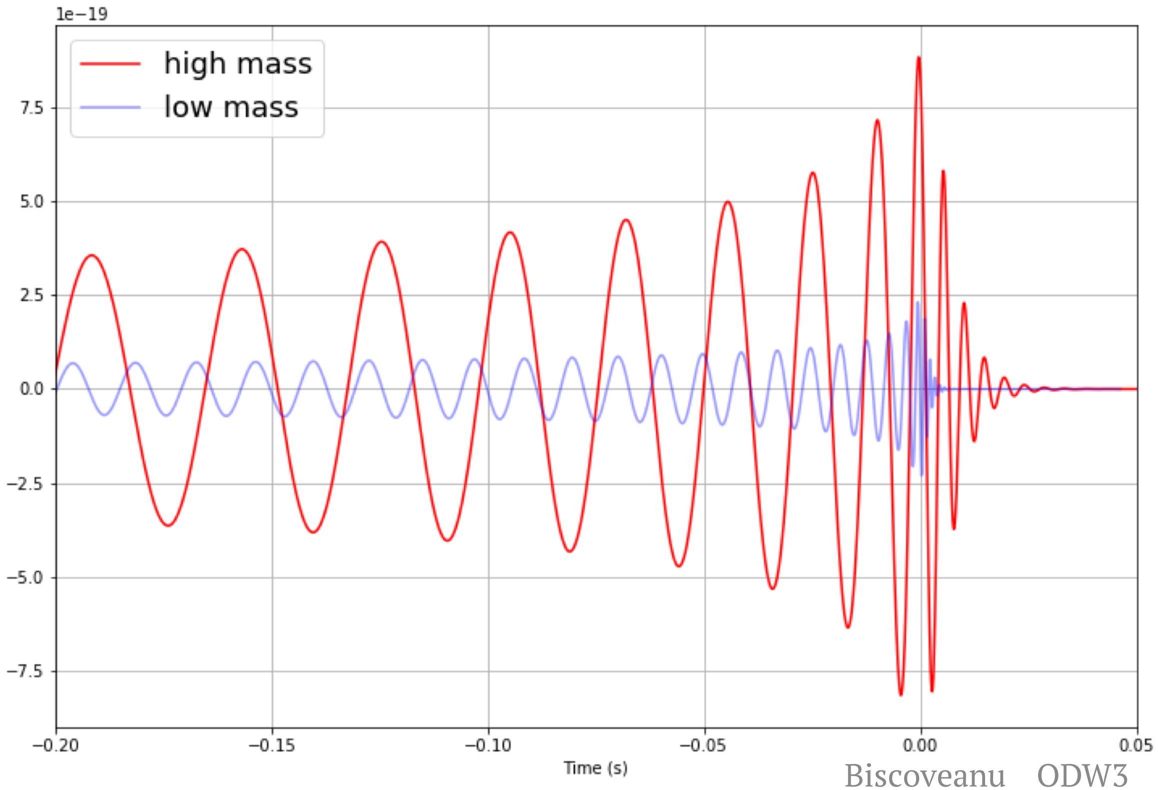
$$\tilde{h}_\times(f) = \mathcal{A}_{\mathrm{GW}}(f)\cos\iota\sin\phi_{\mathrm{GW}}(f)$$

$$\tilde{h}(f) = F_+\tilde{h}_+(f) + F_\times\tilde{h}_\times(f)$$

- Review:
  - Two polarizations
  - Dependence on mass, spins, distance, etc. encoded in amplitude and phase
  - Antenna patterns depend on the detector geometry and encode the effect of the extrinsic parameters

Effect of Mass

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

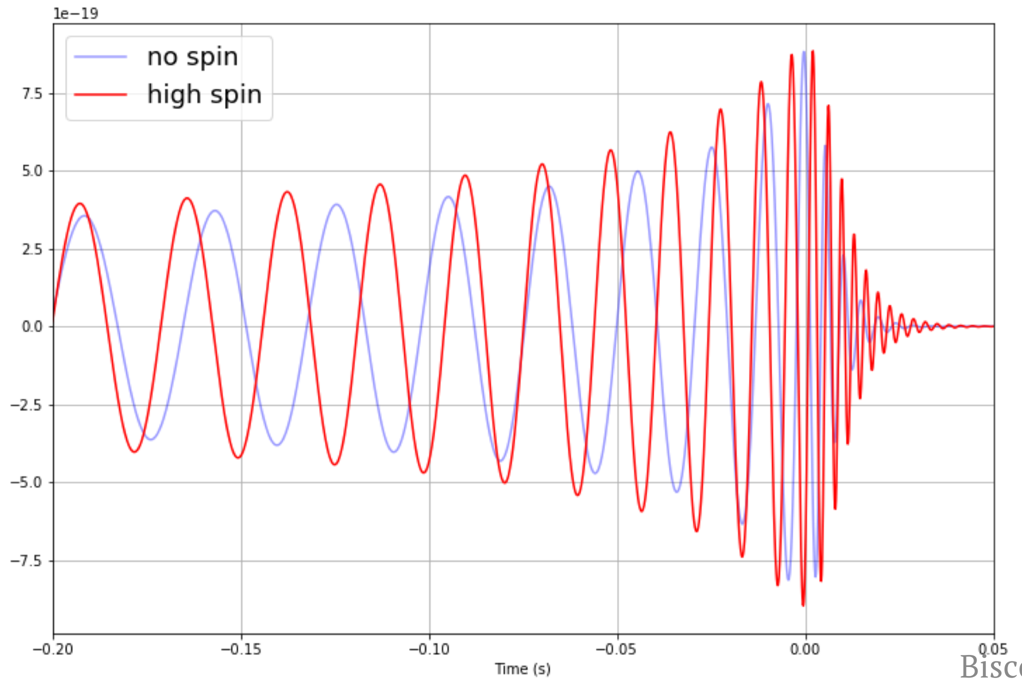


$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

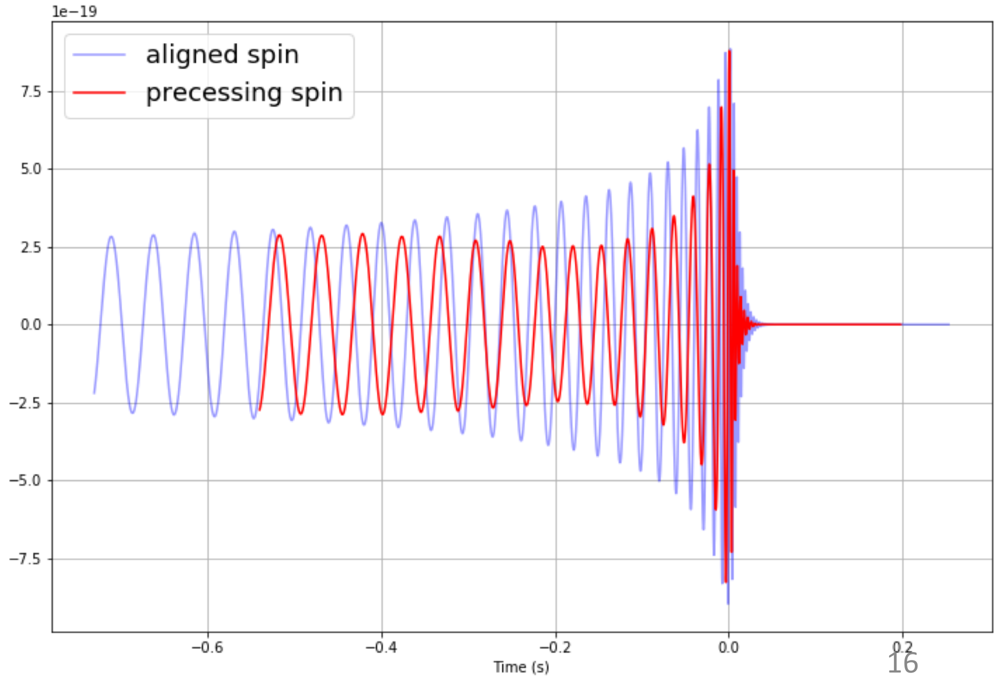
$$\mathcal{A}_{\mathrm{GW}} \propto \frac{\mathcal{M}^{5/6} f^{-7/6}}{d_L}$$

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
- 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 ($\propto d_L^2$)

Bayesian Model Selection

- Simple example – signal versus noise
- Noise evidence is the likelihood evaluated in with no signal model

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

- Bayes factor:

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

Bayesian Model Selection

- Another example – aligned vs precessing spins

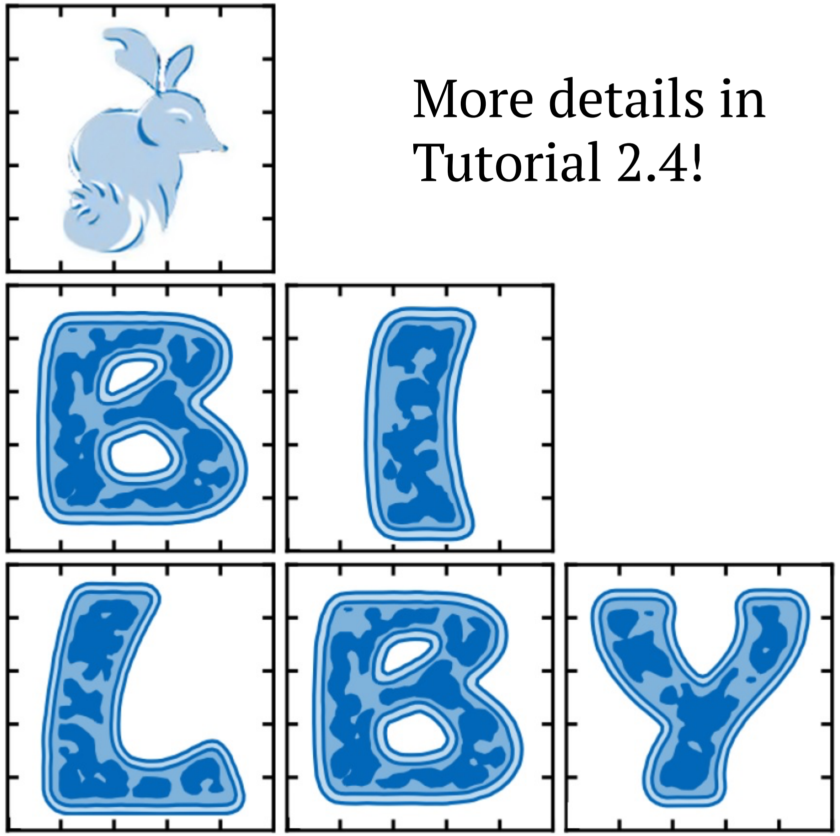
$$\mathrm{BF}_P^A = \frac{\int \mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|A)\ d\theta}{\int \mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|P)\ d\theta}$$

- 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

Bilby



More details in Tutorial 2.4!

- The **B**ayesian **I**nference **L**i**b**rar**y** 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

Hierarchical Modeling

- What if we want to study a population of sources?
- Example: what is the distribution of primary masses for binary black holes?
- The model for the population distribution is called the hyper-prior: $\pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})$ where $\boldsymbol{\theta}$ are the original parameters, and $\boldsymbol{\Lambda}$ are the hyper-parameters
- Example: parameterize the primary mass distribution as a power law

$$\pi(m_1|\alpha) \propto m_1^{\alpha}$$

Hierarchical Modeling

- The new likelihood is the original likelihood marginalized over the original parameters:

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

Hyper-prior

Original likelihood (doesn't depend on hyper-parameters)

$$= \int d\boldsymbol{\theta} \mathcal{L}(d|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})$$

Original evidence

$$= \int d\boldsymbol{\theta} \frac{p(\boldsymbol{\theta}|d)\mathcal{Z}_\theta}{\pi_0(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\boldsymbol{\Lambda})$$

Original prior

Combining multiple events

- Hyper-parameter likelihood for a population is the product of individual-event likelihoods:

$$\mathcal{L}(\{d\}|\mathbf{\Lambda}) = \prod_{j}^{N} \mathcal{L}(d_j|\mathbf{\Lambda})$$

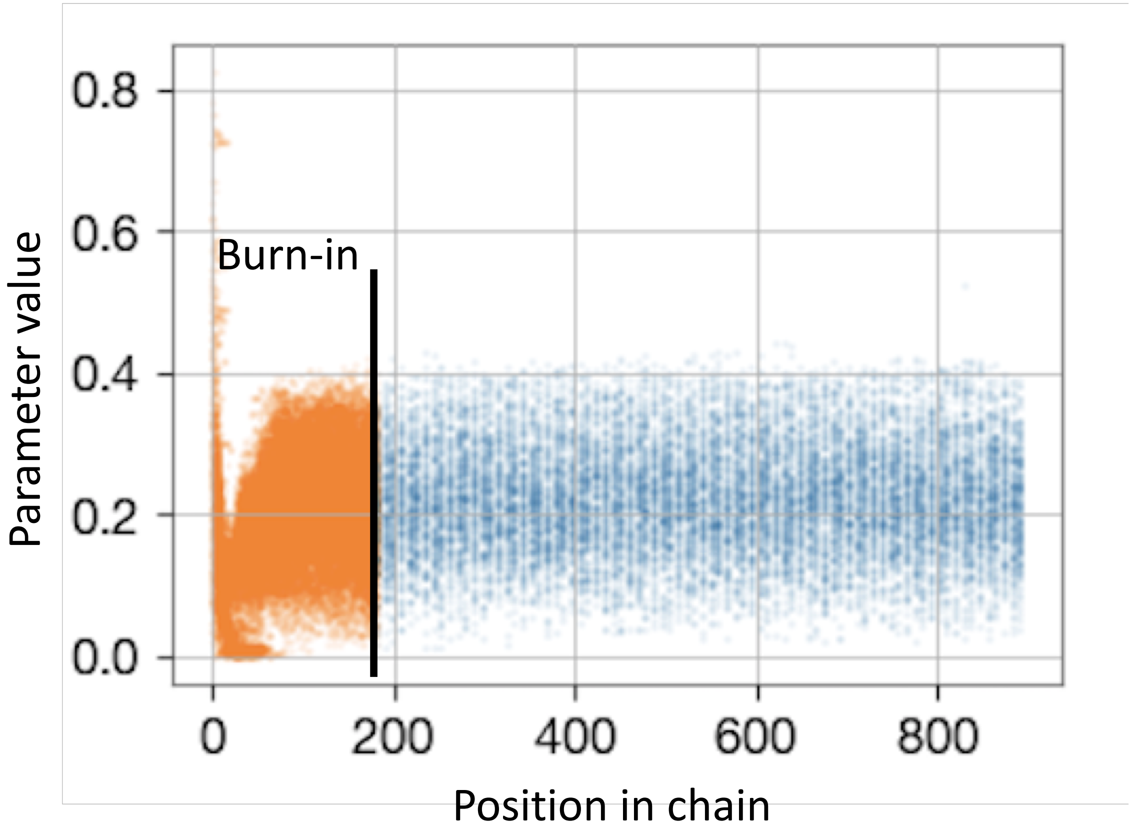
- Complications arise due to selection biases – more likely to detect more massive systems that are close by
- Need to account for probability of detecting signals across the parameter space of interest

Additional Resources

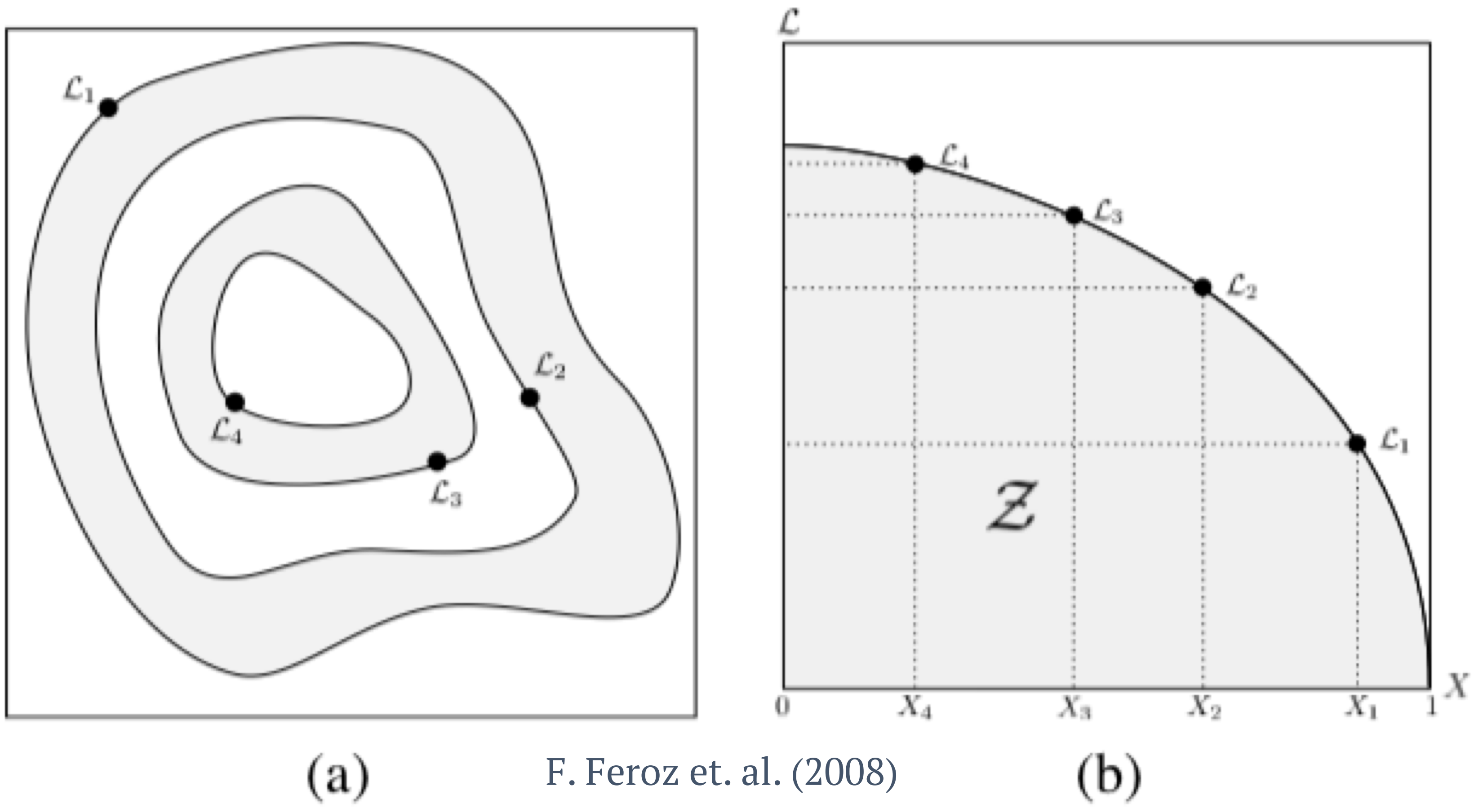
- https://lscsoft.docs.ligo.org/bilby/ - Bilby documentation
- https://chi-feng.github.io/mcmc-demo/ - cool animations of MCMC
- Further reading:
  - Veitch et. al. (2015) https://arxiv.org/pdf/1409.7215.pdf
  - Ashton et. al. (2018) https://arxiv.org/abs/1811.02042
  - Thrane and Talbot (2019) https://arxiv.org/pdf/1809.02293.pdf

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
- Determining a suitable proposal density function is the hard part of sampling – a simple example is a Gaussian centered on the current location
- 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



Nested Sampling



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

Using discrete posterior samples

- For a probability distribution represented by a discrete set of posterior samples:

$$E[f] = \int f(x)p(x)dx$$

$$= \frac{1}{N}\sum_{i}^{N} f(x_i)$$

- So the hyper-PE likelihood becomes a ratio of priors:

$$\mathcal{L}(d|\mathbf{\Lambda}) = \frac{\mathcal{Z}_\theta}{n}\sum_{i}^{n}\frac{\pi(\boldsymbol{\theta}_i|\mathbf{\Lambda})}{\pi_0(\boldsymbol{\theta}_i)}$$