Data analysis techniques
Part II

Frédérique Marion

1st MaNiTou Summer School on Gravitational Waves

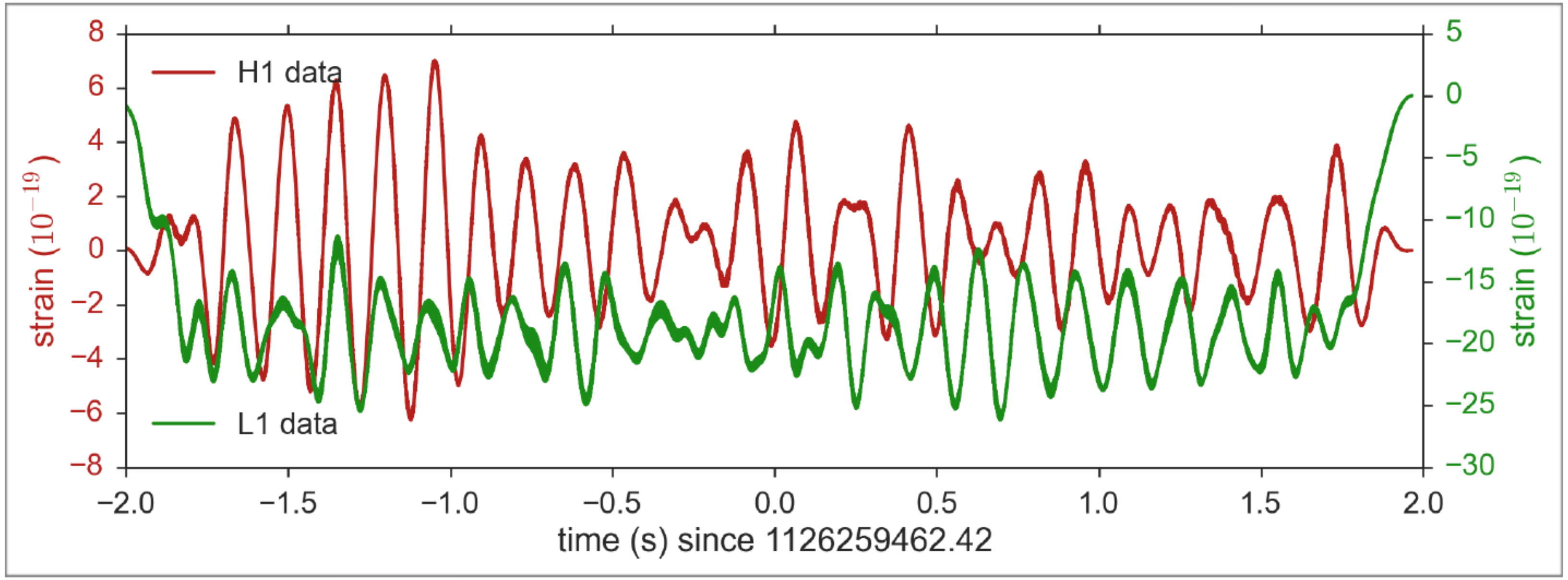
July 2022

Scope

Extracting the science, mostly through Bayesian analyses of

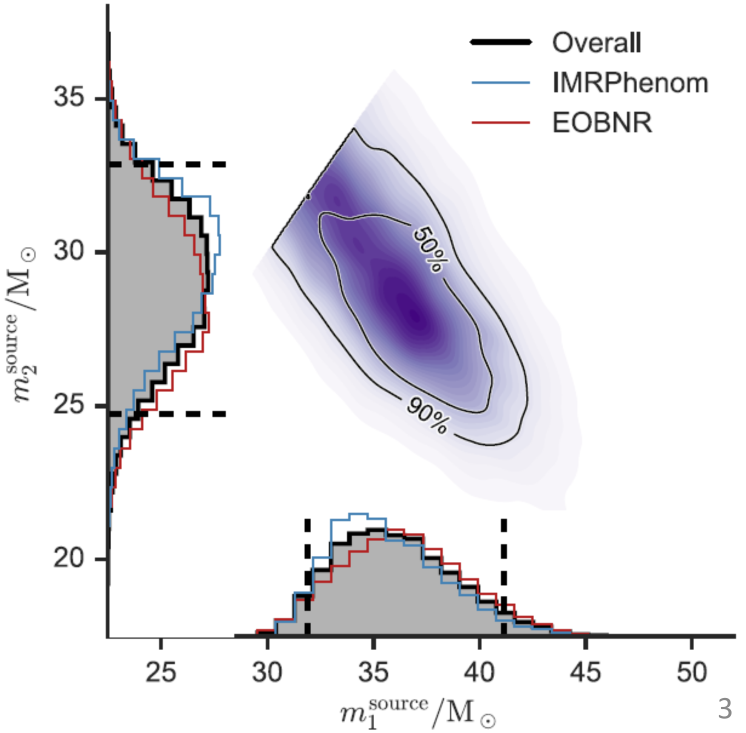
- Individual events
- Collections of events

From data to astrophysical parameters



- GW150914
  - September 14, 2015 at 09:50:45 UTC

Source-frame primary mass $m_1^{\rm source}/M_\odot$	$35.8^{+5.3\pm0.9}_{-3.9\pm0.1}$
Source-frame secondary mass $m_2^{\rm source}/M_\odot$	$29.1^{+3.8\pm0.1}_{-4.3\pm0.7}$



Recommended reading:
Parameter Estimation with Gravitational Waves
Christensen & Meyer, arXiv:2204.04449

Parameter estimation via Bayesian inference

- Assume data **d** are described by model *M* with parameters $\vec{\theta}$
- Use Bayes' theorem to infer posterior probability distribution for parameters $\vec{\theta}$, given data **d**

likelihood

$$p(\vec{\theta}\,|\mathbf{d}, M) = \frac{p(\mathbf{d}|\,\vec{\theta}\,, M)\,p(\,\vec{\theta}\,|M)}{p(\mathbf{d}|M)}$$

prior
a priori knowledge about $\vec{\theta}$

posterior
a posteriori knowledge about $\vec{\theta}$

evidence

Model for the data

$$d = R[h] + n$$

available data

detector response to GW

detector noise

Assumption
- ✓ data properly calibrated

Assumptions
- ✓ Gaussian
- ✓ Stationary
- ✓ Uncorrelated across detectors

Likelihood

$$p(\mathbf{d}|\vec{\theta}, M)$$

$$d = R[h] + n$$

- Noise probability distribution $p(n)$
- How likely is the residual $d - R[h]$ assuming it is noise?
  - Probability of drawing the residual from the noise distribution
  - $p(n)$ ➔ $p(d - R[h]) \equiv p(d|\vec{\theta})$
  - Once we have a signal model, the noise model defines the likelihood

Noise model

- Gaussian noise
  - Single data point $p(n_i) \propto e^{-n_i^2/2\sigma^2}$
  - Multiple data points $p(n_1, n_2 \dots n_N) \propto e^{-\frac{1}{2}\Sigma\ n_i C_{ij}^{-1} n_j}$
- Stationary noise

Time domain

C_{ij} only depends on $(j-i)$

$$C_{i(i+m)} = C(m) = E[n_i n_{i+m}] = \frac{1}{N}\Sigma_i n_i n_{i+m}$$

$$C(\tau) = E[n(t)n(t+\tau)] = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{+T/2} n(t)n(t+\tau)dt$$

Power spectrum is Fourier transform of autocorrelation

Frequency domain

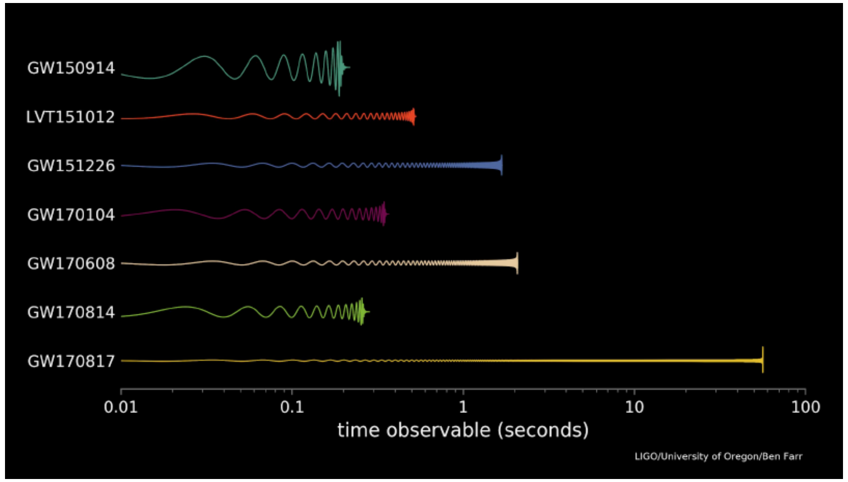
$$< \tilde{n}(f)\tilde{n}^*(f') >= \frac{1}{2}S_n(|f|)\delta(f-f')$$

$\tilde{n}_i \tilde{C}_{ij}^{-1}\tilde{n}_j \quad \tilde{C}_{ij} = \frac{1}{2}S_n(f_i)\delta_{ij}$

$$4\int_0^\infty \frac{\tilde{n}(f)\tilde{n}^*(f)}{S_n(f)}df \equiv (n|n)$$

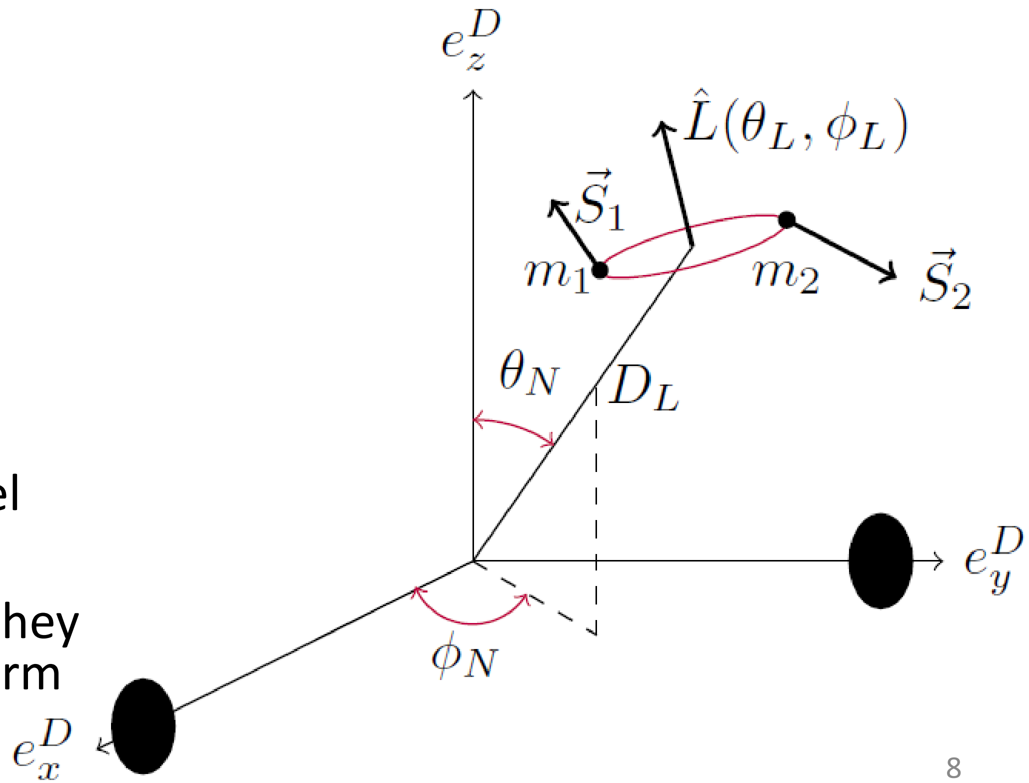
$$p(\tilde{n}_1, \tilde{n}_2 \dots \tilde{n}_N) \propto e^{-\frac{1}{2}(n|n)}$$

Signal model



- In general, compact binary is described by up to 19 parameters
  - Intrinsic parameters drive system dynamics
    - Masses (2)
    - Spins (6)
    - Deformability for neutron stars (2)
    - Eccentricity (2)
  - Extrinsic parameters impact measured signal
    - Position : luminosity distance, right ascension, declination (3)
    - Orientation: inclination, polarization (2)
    - Time and phase at coalescence (2)

- Reliable waveform models exist
  - Not all physical effects are accounted for in any given model
  - Computing time is an issue for parameter estimation
  - Various models used, differing both in the physical effects they describe and the methods they use to compute the waveform



Alternative signal model

- Strain waveform reconstructed with minimal-assumption signal model
  - Linear combination of elliptically polarized sine-Gaussian wavelets
  - Algorithm varying both model parameters and model dimension

$$\Psi(t; A, f_0, Q, t_0, \phi_0) = A e^{-(t-t_0)^2/\tau^2} \cos(2\pi f_0 (t - t_0) + \phi_0) \qquad \tau = Q/(2\pi f_0)$$

$$h_+(f) = \sum_{j=0}^{\boxed{N_s}} \Psi(f; \boxed{A_j, f_{0j}, Q_j, t_{0j}, \phi_{0j}}) \qquad h_\times = \boxed{\epsilon} h_+ e^{i\pi/2}$$

$$(\mathbf{R} \star \mathbf{h})_i(f) = \left(F_i^+(\theta, \phi, \psi) h_+(f) + F_i^\times(\boxed{\theta, \phi, \psi}) h_\times(f) \right) e^{2\pi i f \Delta t_i(\theta,\phi)}$$

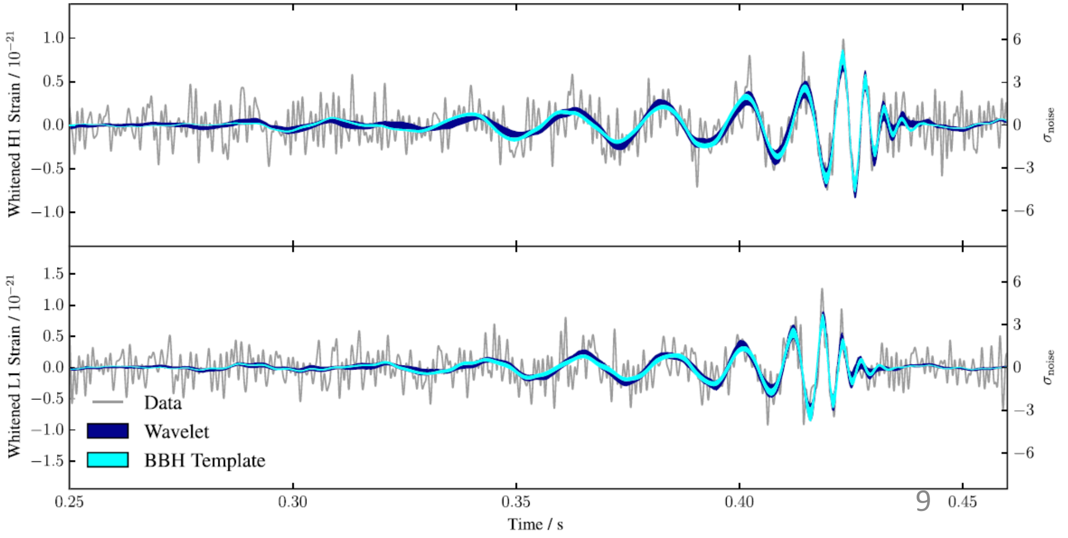
model dimension

parameters for each wavelet

common extrinsic parameters

GW150914

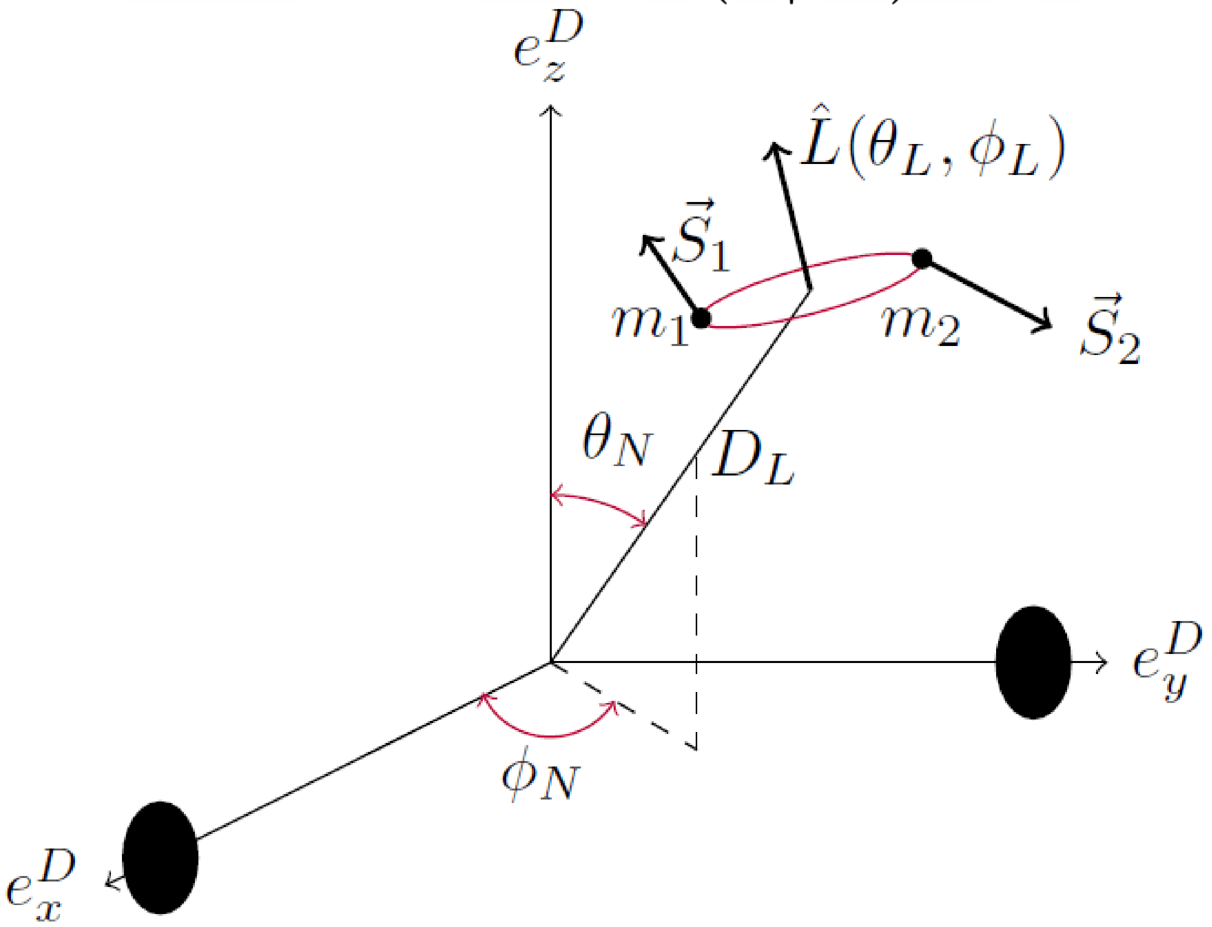
PRL 116, 241102 (2016)



BayesWave – Cornish & Littenberg *Class. Quantum Grav.* **32** 135012 (2015)

Priors: extrinsic parameters

$$p(\vec{\theta}|\mathbf{d}, M) = \frac{p(\mathbf{d}|\vec{\theta}, M)p(\vec{\theta}|M)}{p(\mathbf{d}|M)}$$

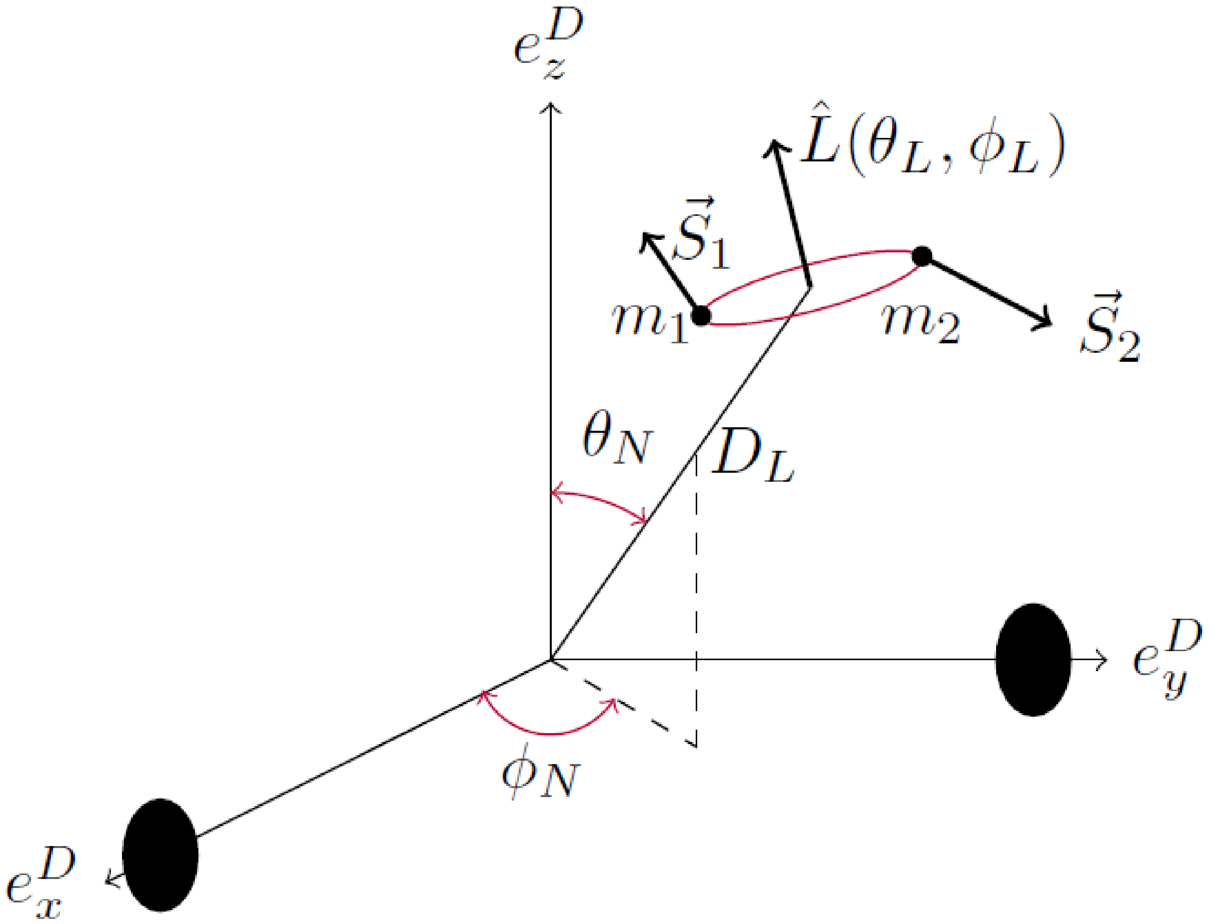


No unique choice of priors!

D_L	Uniform in volume
θ_N, ϕ_N	Uniform in the sky
θ_L, ϕ_L	Uniform in direction

Priors: intrinsic parameters

$$p(\overrightarrow{\theta}|\mathbf{d}, M) = \frac{p(\mathbf{d}|\overrightarrow{\theta}, M)p(\overrightarrow{\theta}|M)}{p(\mathbf{d}|M)}$$



No unique choice of priors!

m_1 m_2	Uniform in some range
$\vec{S}_1$ $\vec{S}_2$	Uniform in direction Magnitude uniform in $[0, Gm_i^2/c]$
Λ_1 Λ_2	Uniform in (0, 5000)

Evidence

$$p(\vec{\theta}|\boldsymbol{d}, M) = \frac{p(\boldsymbol{d}|\vec{\theta}, M)p(\vec{\theta}|M)}{p(\boldsymbol{d}|M)}$$

- Unimportant normalization factor for parameter estimation
  - Evidence = marginal likelihood

$$p(\boldsymbol{d}|M) = \int_{\Omega_{\vec{\theta}}} p(\boldsymbol{d}|\vec{\theta}, M)p(\vec{\theta}|M)d\vec{\theta}$$

  - Computation typically difficult
    - Sometimes built-in in sampling algorithm, e.g. nested sampling
- Important for model selection

Evidence and model section

- Posterior odds ratio of model M_1 vs model M_2

$$\mathcal{O}_{12} = \frac{p(M_1|\boldsymbol{d})}{p(M_2|\boldsymbol{d})} = \frac{p(M_1)p(\mathbf{d}|M_1)}{p(M_2)p(\mathbf{d}|M_2)}$$

Bayes factor

- Ratio of posterior to prior odds
  - Ratio of the evidences

$$\mathcal{B}_{12} = \frac{p(M_1|\boldsymbol{d})}{p(M_1)}\frac{p(M_2)}{p(M_2|\boldsymbol{d})}$$

$$= \frac{\int_{\Omega_{\vec{\theta_1}}} p(\mathbf{d}|\vec{\theta_1}, M_1)p(\vec{\theta_1}|M_1)d\vec{\theta_1}}{\int_{\Omega_{\vec{\theta_2}}} p(\mathbf{d}|\vec{\theta_2}, M_2)p(\vec{\theta_2}|M_2)d\vec{\theta_2}}$$

M_1 = signal + Gaussian noise

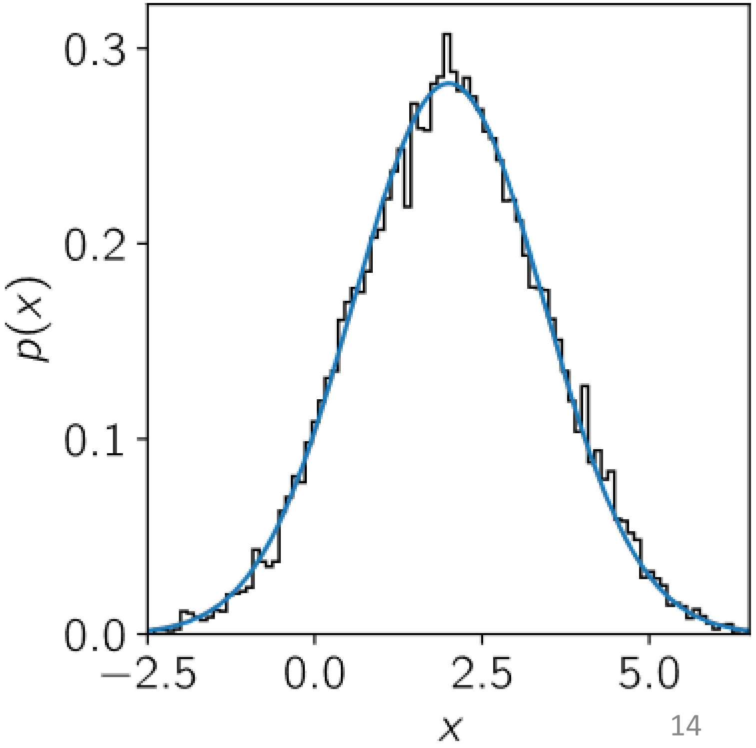
M_2 = Gaussian noise

$$\ln \mathcal{B}_{s/n} \sim \frac{1}{2}\mathrm{SNR}^2$$

Sampling the posterior

- Algorithm needed to explore multi-dimensional parameter space
  - Cost of brute-force method – compute posterior pdf on fine grid – not prohibitive only for very low dimensions
    - Most general model for CBC source has 19 parameters!
  - Efficient stochastic sampling algorithm needed
- Sampling: set of (n-dim) parameter values that together give a fair representation of the posterior pdf
- Markov-chain Monte Carlo (MCMC) algorithms generate samples iteratively, via biased random walk through parameter space
  - Walk based on two rules
    - How to draw new position from current position
    - How to decide whether to accept new sample or repeat previous one
      - Involves likelihood and prior values
  - Possibly using parallel chains
  - Need (empirical) ways to check convergence of sample chain
- Many different samplers – LVK use several of them – with different efficiency, ability to deal with multi-modality, etc.
  - e.g. nested sampling

Recommended reading:
Data Analysis Recipes: Using Markov Chain Monte Carlo
Hogg & Foreman-Mackey
ApJS **236** 11 (2018)



Calibration

- PE needs to take into account that calibration is not perfect

$$\tilde{h}_{\rm obs}(f) = \tilde{h}(f)\,(1+\delta A(f))\exp\left(i\delta\phi(f)\right)$$

  - Model amplitude and phase errors as cubic splines

$$\delta A(f) = p_s\left(f; \{f_i, \delta A_i\}\right)$$

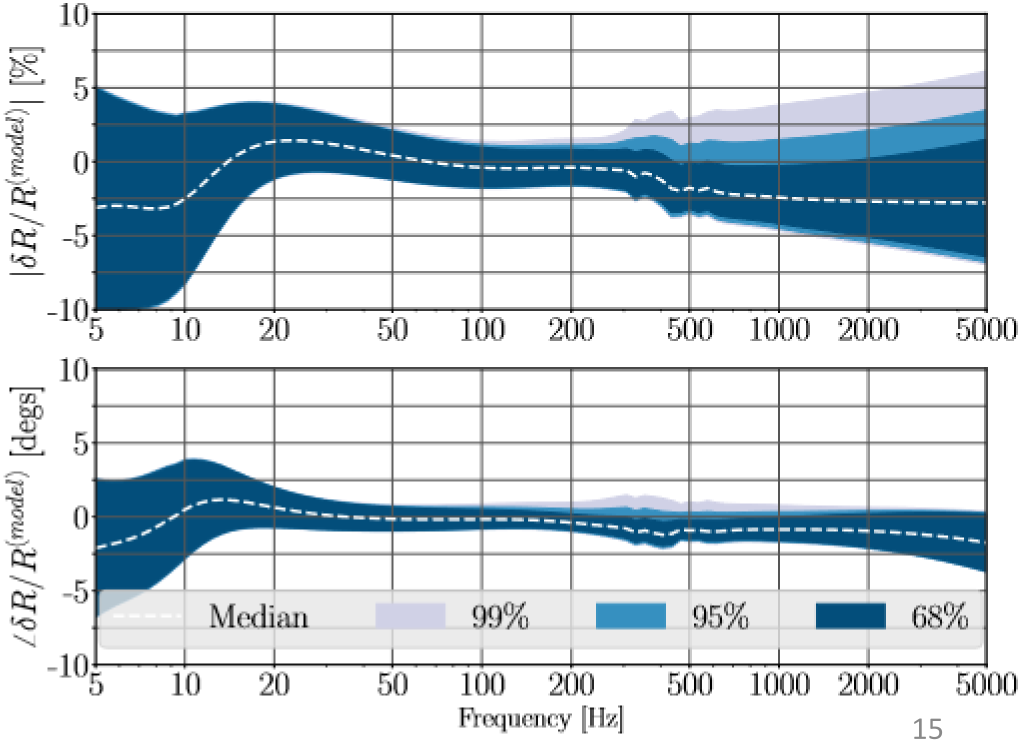
$$\delta\psi(f) = p_s\left(f; \{f_i, \delta\phi_i\}\right)$$

  - Priors on parameters informed by calibration uncertainties

$$p\left(\delta A_i\right) = N(0, \sigma_A) \qquad p\left(\delta\psi_i\right) = N(0, \sigma_\psi)$$

  - PE results marginalized over calibration parameters

Example: L1 calibration uncertainties during O2 run

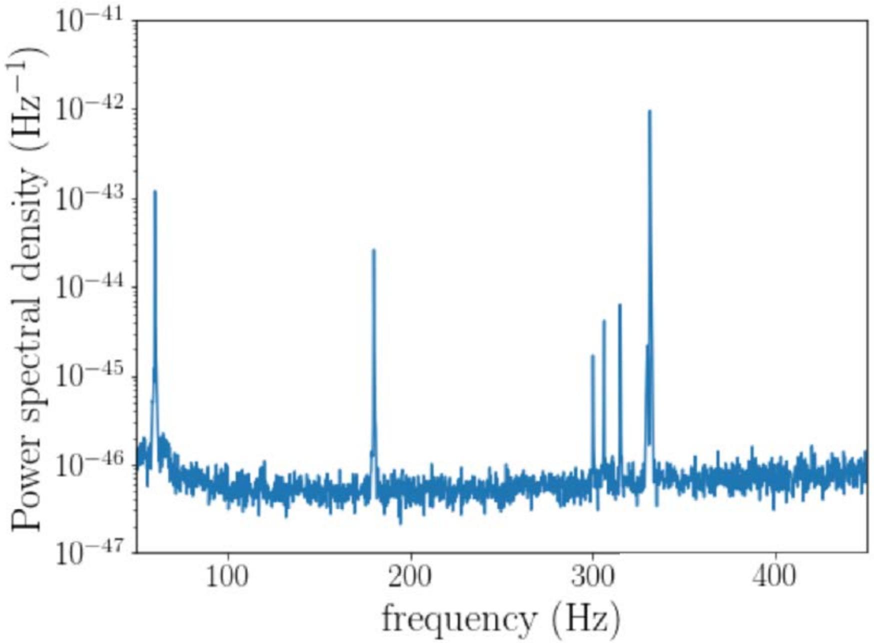


Cahilane+ Phys. Rev. D 96, 102001 (2017)

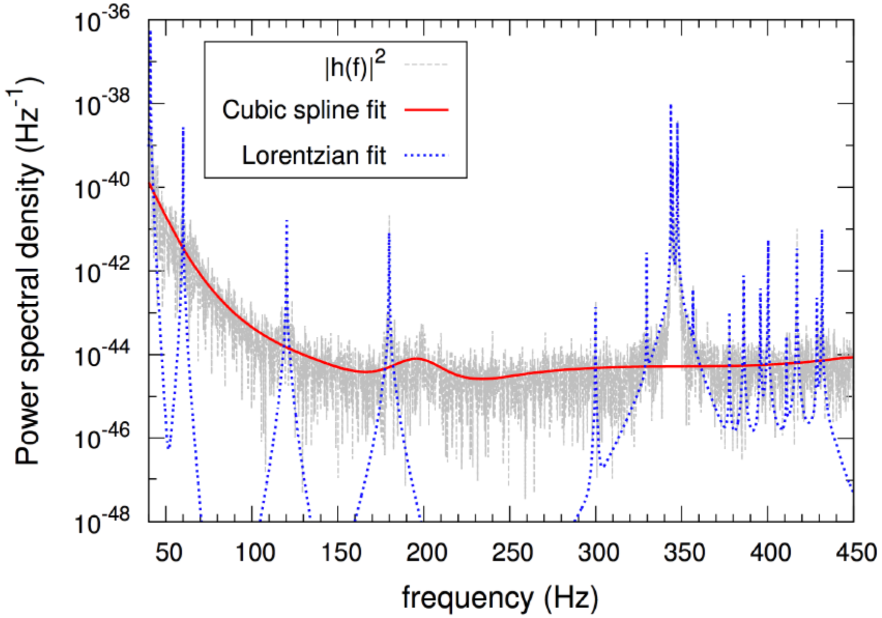
Noise model: spectrum

- ❑ Detector noise is not stationary on long time scales
- ❑ Locally, stationarity assumption is reasonable if using a locally representative spectrum

Off source estimate



On source estimate



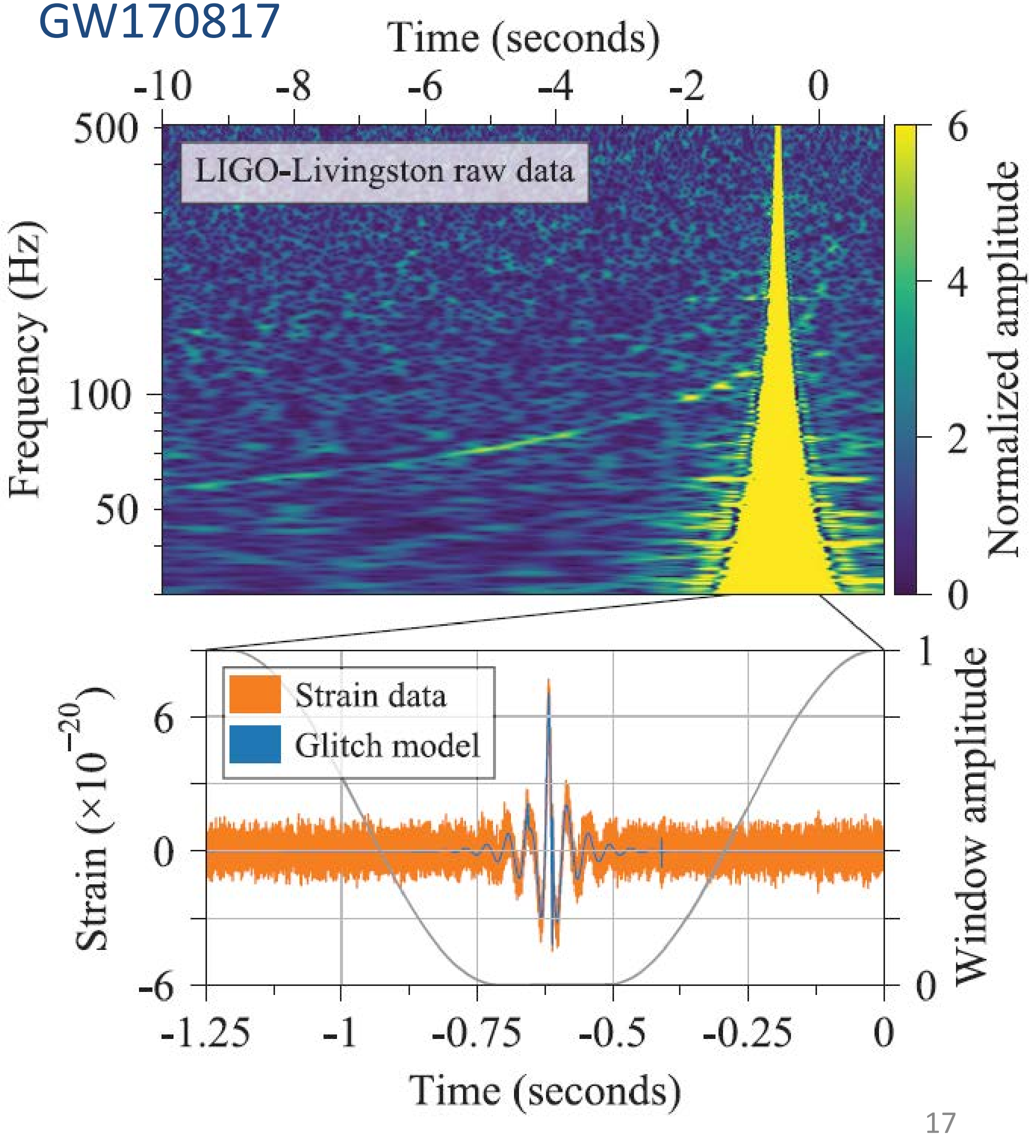
BayesLine – Littenberg & Cornish Phys. Rev. D 91, 084034 (2015)

Noise model: glitch removal

- Detector noise is not Gaussian on long time scales
- Locally, Gaussian assumption is reasonable provided data are free of excess noise – aka glitches
  - If glitch present, include in model or remove from data

$$d = R[h] + n + g$$

GW170817

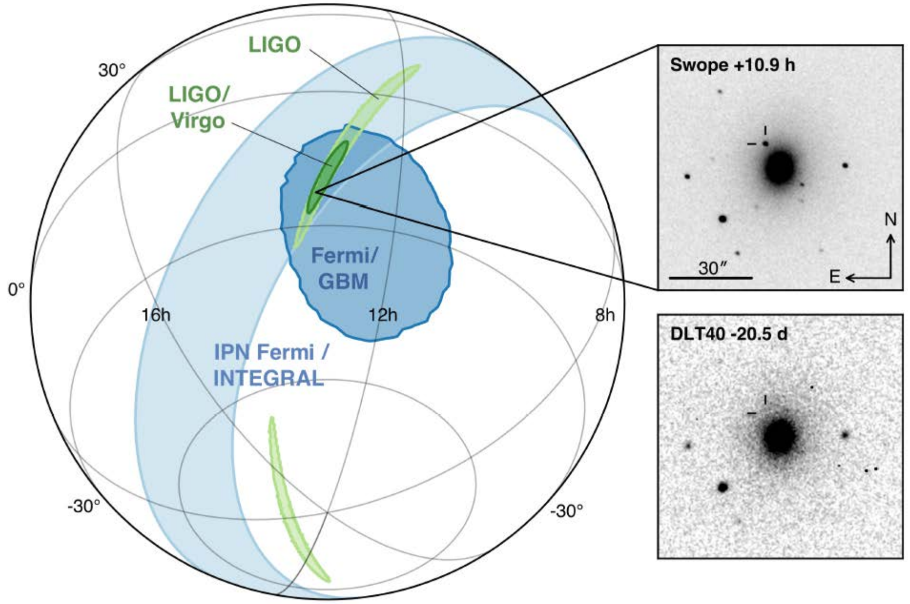


PRL 119, 161101 (2017)

Rapid parameter estimation

- Parameter estimation requires long computing times
  - A few hours for short BBH signals
  - Weeks for BNS signals
  - Driven by evaluating likelihood (inc. computing waveform) at each step
- Various strategies to reduce computational cost
  - Waveform acceleration
  - Parallelization
- Low-latency localization of sources for electromagnetic follow-up
  - Focus is on extrinsic parameters
    - Fix intrinsic parameters to values reported by search pipelines
  - Information crucial for localization is encapsulated in matched-filter estimates of times, amplitudes, and phases on arrival at the detectors
  - Compute posterior distribution of extrinsic parameters, provide (good!) approximate marginal posterior distribution of sky location within minutes

GW170817



ApJL 848 L12 (2017)

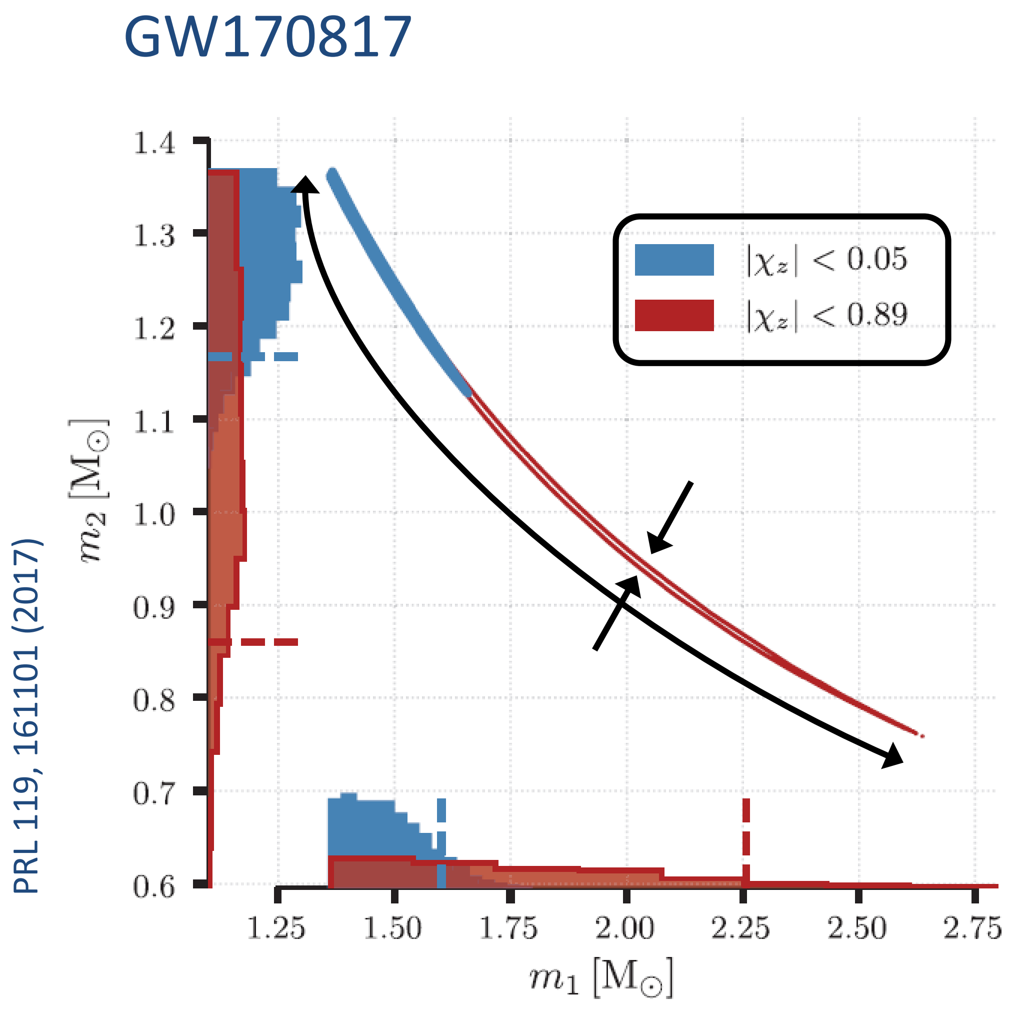
BAYESTAR – Singer & Price Phys. Rev. D 93, 024013 (2016)

Presenting and quoting results

- Multi-dimensional posterior samples are end result of inference
  - Contain all information ➔ Release full set of posterior samples
  - Not easily digestible
  - We want 1D and/or 2D plots and summary statistics
  - We need to quote statistical uncertainties
  - We need to quote systematic uncertainties

Recommended reading for LVK members:
Quoting parameter-estimation results
Berry et al., LIGO-T1500597 (2015)

Corner plots



- Choose a pair of parameters, draw 2D and 1D posteriors marginalized over all other parameters

$$p(m_1, m_2|\boldsymbol{d}) = \int_{\vec{\theta}_{\text{other}}} p(\vec{\theta}_{\text{other}}, m_1, m_2|\boldsymbol{d})\, d\vec{\theta}_{\text{other}}$$

  - Highlights parameter correlations

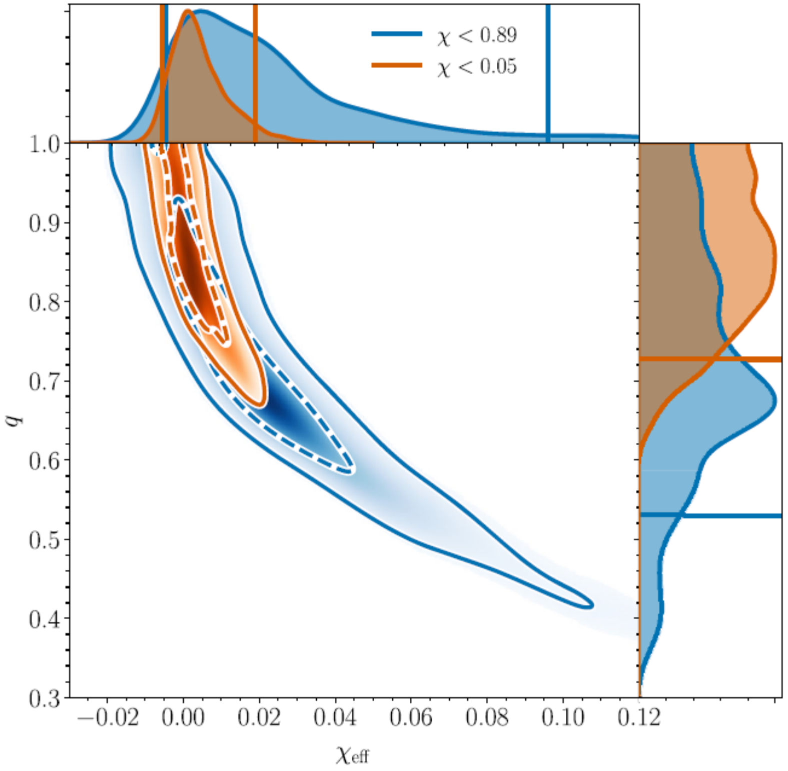
The chirp mass $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ drives the inspiral and is measured very well

The mass ratio $q = m_2/m_1$ enters at higher order and is measured less well

The mass ratio is correlated with the spin

Corner plots (cont.)

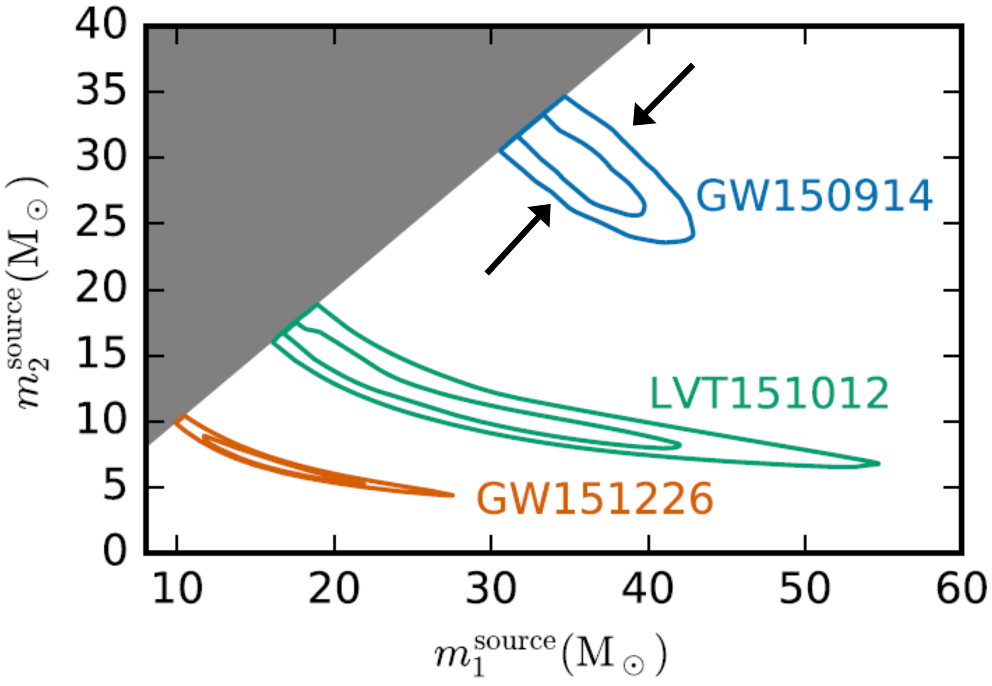
GW170817



PRX 9, 011001 (2019)

$$\chi_{\rm eff} = \frac{m_1 \chi_{1z} + m_2 \chi_{2z}}{m_1 + m_2}$$

- For high-mass systems, merger-ringdown is a significant part of the signal, driven by the total mass

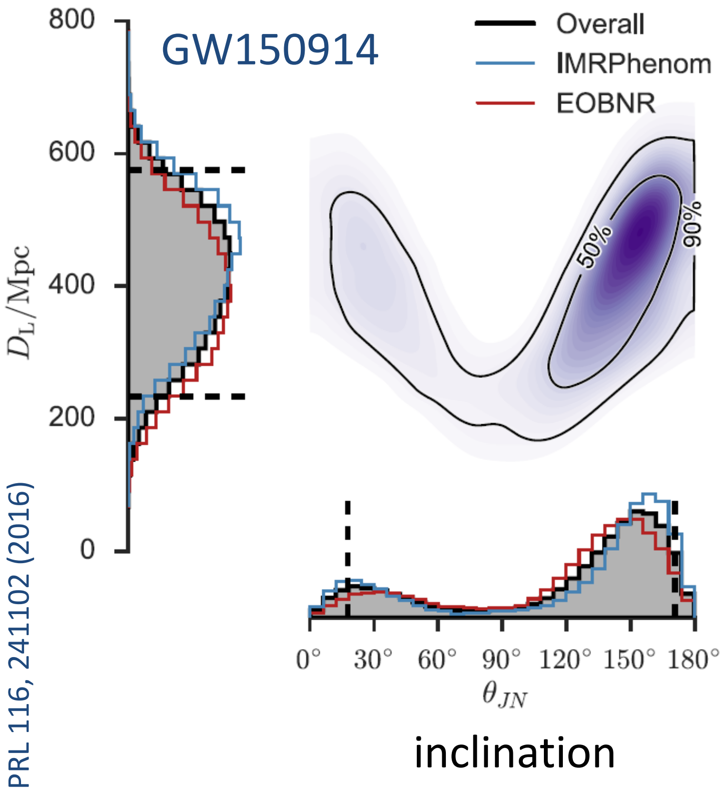


PRX 6, 041015 (2016)

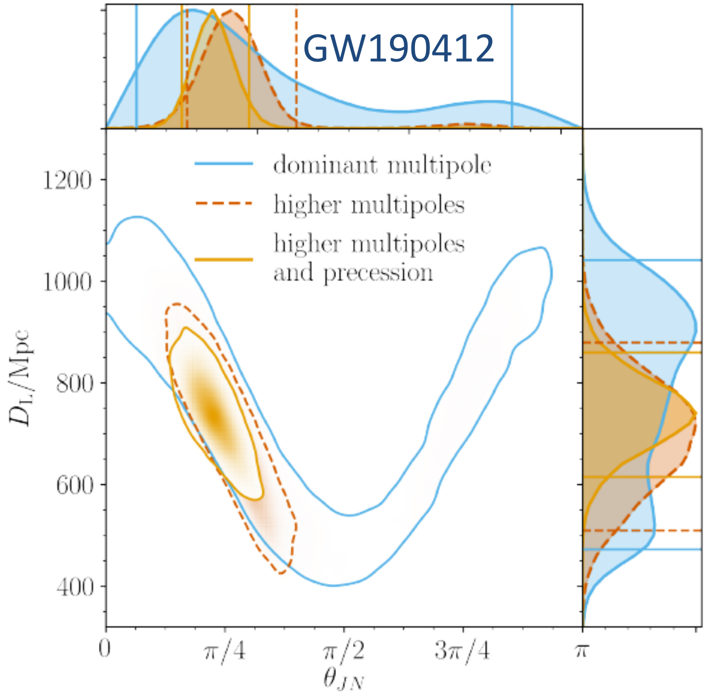
Corner plots (cont.)

- From GW signal, difficult to distinguish distant, well-oriented source from nearby, ill-oriented source
  - Correlation between luminosity distance and inclination (and direction)

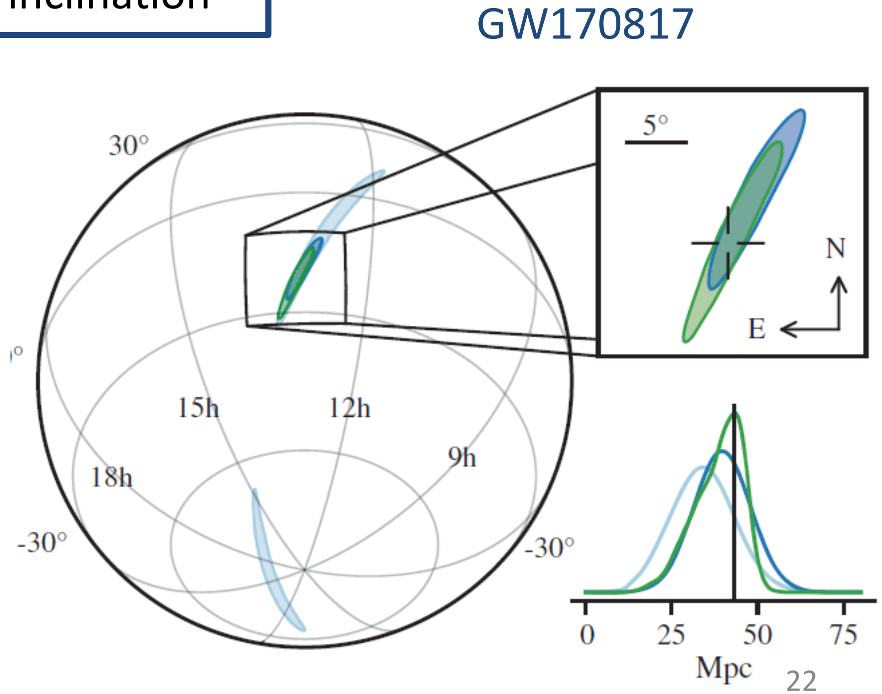
- Asymmetric system GW190412 (30+8 $M_\odot$)
  - Presence of higher-order modes helps lifting degeneracy between distance and inclination



PRL 116, 241102 (2016)



PRD 102, 043015 (2020)



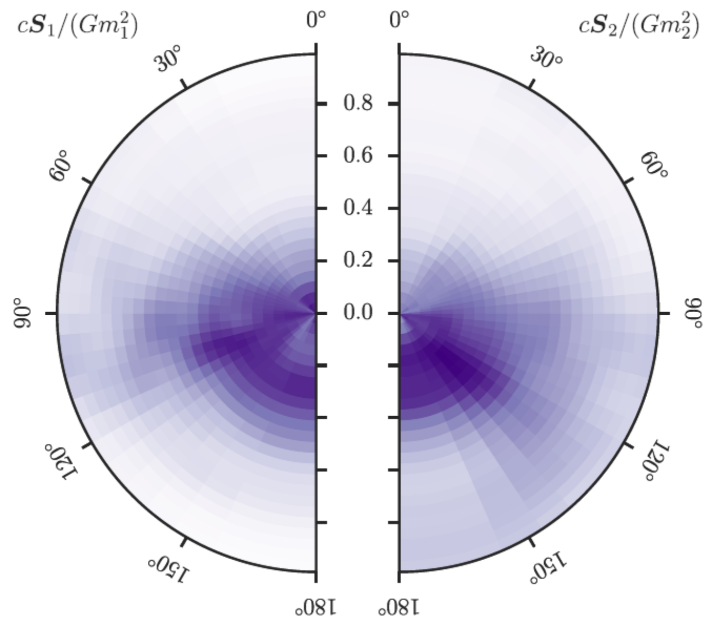
PRL 119, 161101 (2017)s

Spins: disk plots

- Spins enter at higher order in system dynamics and have subtle effects on GW waveform
  - Difficult to measure
  - Unless precession changes inclination over time and induces spectacular amplitude and phase modulation

GW150914

spin aligned with orbital angular momentum

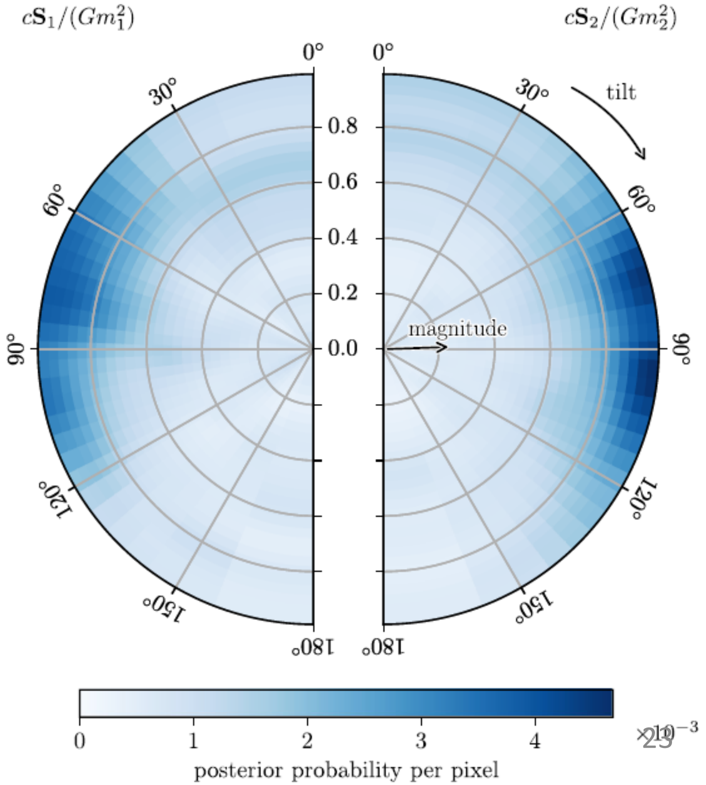


spin anti-aligned with orbital angular momentum

PRL 116, 241102 (2016)

- 2D posterior probability for tilt angle and spin magnitude for each object
- Tiles constructed linearly in spin magnitude and cosine of tilt angle (identical prior probability)
- Color indicates posterior probability per pixel, marginalized over azimuthal angle

GW190521



PRL 125, 101102 (2020)

Best estimates

- Maximum likelihood (ML)
  - Point where model best fits data
  - Ignores prior information

- Maximum posterior (maximum a posteriori, MAP)
  - Peak of posterior probability distribution – modal value, most probable point
  - Ambiguous definition: global maximum or maximum of each 1D distribution?
  - Not invariant under reparametrization
  - Not necessarily a typical value, not very useful for multimodal distributions

- Posterior mean
  - Expectation value of distribution
  - Better traces position of posterior mass than MAP (= MAP for Gaussian distribution)
  - Not invariant under reparametrization – Not sensible to combine means for different parameters
  - Not necessarily coincides with probable posterior value – e.g. for bimodal distribution

- Posterior median
  - Position of 50% quantile
  - Gives good indication of position of posterior probability mass
  - Less influenced by tails of distribution than posterior mean
  - Not necessarily coincides with probable posterior value
  - Invariant under monotonic reparametrization – Not sensible to combine medians for different parameters

Statistical uncertainties

- Standard deviation
  - Second moment of distribution
  - Simple interpretation in terms of enclosed probability only for Gaussian distributions
  - Not very useful for skewed or multimodal distributions

- Credible intervals
  - Interval (or volume in n-D) enclosing a given total posterior probability
    - e.g. 90% credible interval covers a total posterior probability of 0.9
  - Can be constructed in multiple ways
  - Choose value for total probability
    - 50% not broad enough
    - 68.269% credible interval ≡ Gaussian 1σ interval, but can be misleading
    - 90% includes most of the potential range
    - 95% ~ Gaussian 2σ interval, but may suffer from inaccurate distribution tails

- Symmetric credible intervals
  - Centered on median, extend outwards such that there is an equal probability in each tail of the distribution
    - e.g. 90% symmetric credible interval: lower bound @ 5% quantile, upper bound @ 95% quantile
  - Natural complement to quoting posterior median
    - But can exclude highly probable values if these occur at edges of parameter space

- One-sided credible regions
  - Start from one edge of parameter space and continue until they contain desired probability
    - e.g. 90% one-sided interval: from minimum value to 90% quantile or from maximum value to 10% quantile
  - Applicable for parameters with definite bound
    - e.g. mass ratio, spin magnitude

Systematic uncertainties

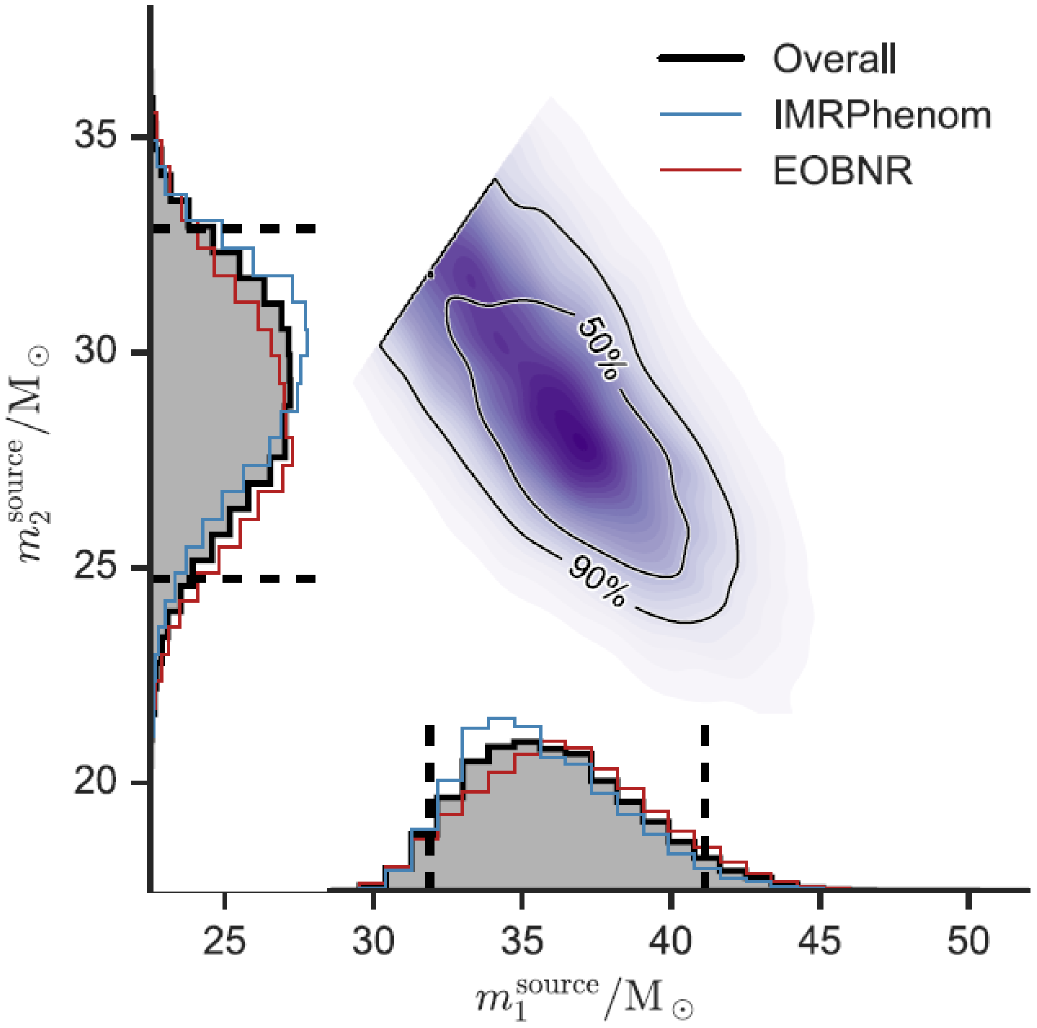
- Compare between results assuming different waveform approximants
  - How to combine posteriors produced with different waveform models?

- Combining ranges
  - Quote maximum and minimum values of all possible statistical uncertainties as overall uncertainty range
  - Conservative, but no simple statistical interpretation

- Averaging posteriors
  - $\equiv$ Marginalize over model uncertainty
  - Average can use weights based on model evidence and prior, or use equal weights
  - Quote point estimate and uncertainty from averaged posterior X^{+Y}_{-Z}
    - Systematic uncertainty folded in overall uncertainty
  - Works only for subspace of common parameters if models have different numbers of parameters
  - Does not construct an estimate for the typical difference between models

- Comparing posterior estimates
  - Start from best posterior estimate (e.g. approximant-averaged posterior) $X^{+Y \pm y}_{-Z \pm z}$
  - Use scatter across approximants to infer systematic uncertainty

GW150914 example

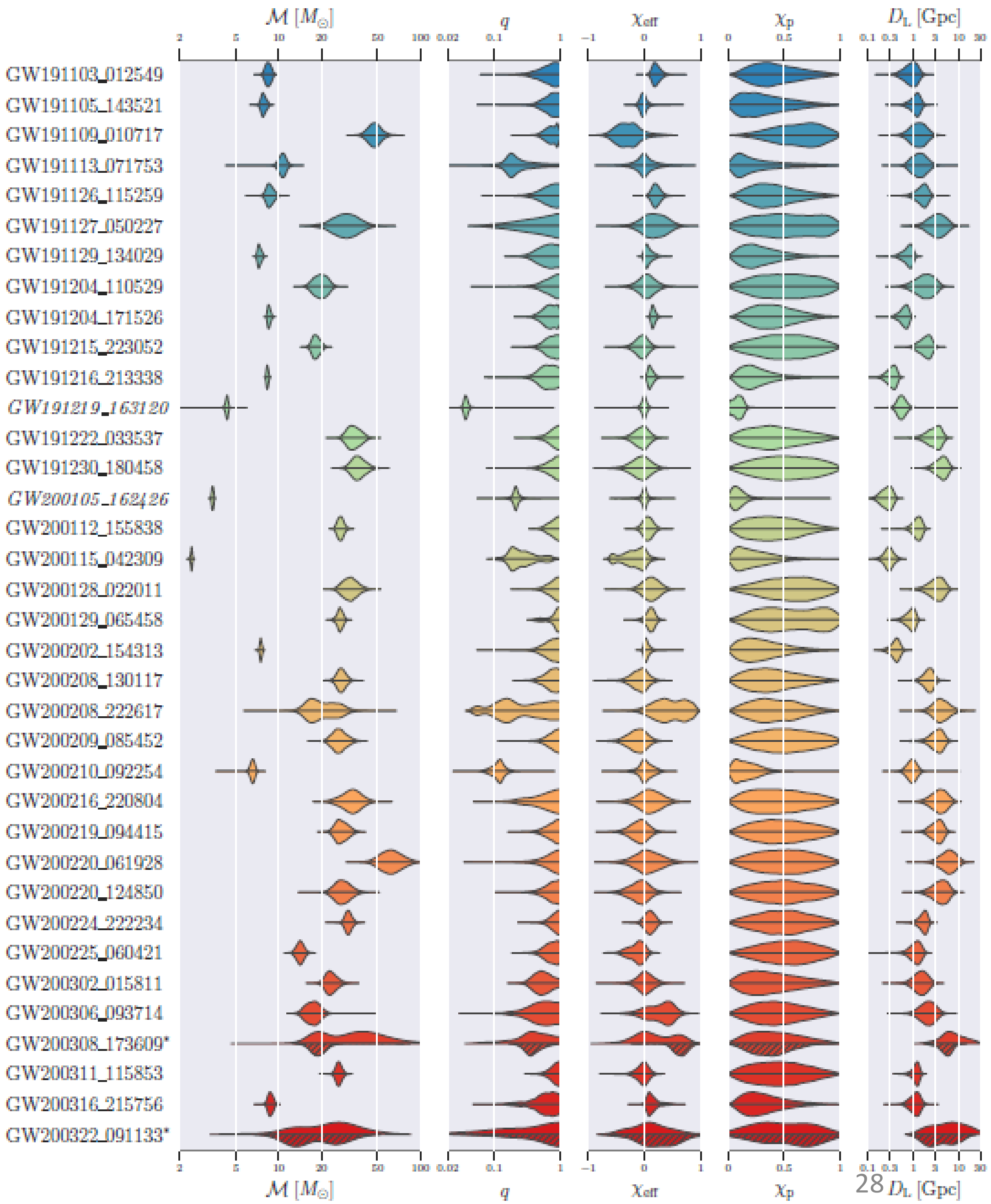


	EOBNR	IMRPhenom	Overall
Source-frame primary mass $m_1^{\rm source}/M_\odot$	$36.3^{+5.3}_{-4.5}$	$35.3^{+5.2}_{-3.4}$	$35.8^{+5.3\pm0.9}_{-3.9\pm0.1}$
Source-frame secondary mass $m_2^{\rm source}/M_\odot$	$28.6^{+4.4}_{-4.2}$	$29.6^{+3.3}_{-4.3}$	$29.1^{+3.8\pm0.1}_{-4.3\pm0.7}$

GWTC3 – arXiv:2111.03606

Multiple events: violin plots

- ❑ Marginal posterior distributions for a selection of parameters for O3b candidates
  - ➢ Color ⇔ date of observation



PE for individual events: Summary

Likelihood

Noise model (Gaussian / deglitched – Stationary / PSD)

Waveform model (GR / generic)

Data calibration

$$p(\vec{\theta}\,|\mathbf{d}, M) = \frac{p(\mathbf{d}|\,\vec{\theta}\,, M)\,p(\,\vec{\theta}\,|M)}{p(\mathbf{d}|M)}$$

Prior

Potentially influential choices

Posterior

Computing time – sampling algorithms

Presenting and quoting digested results

Parameter correlations

Evidence

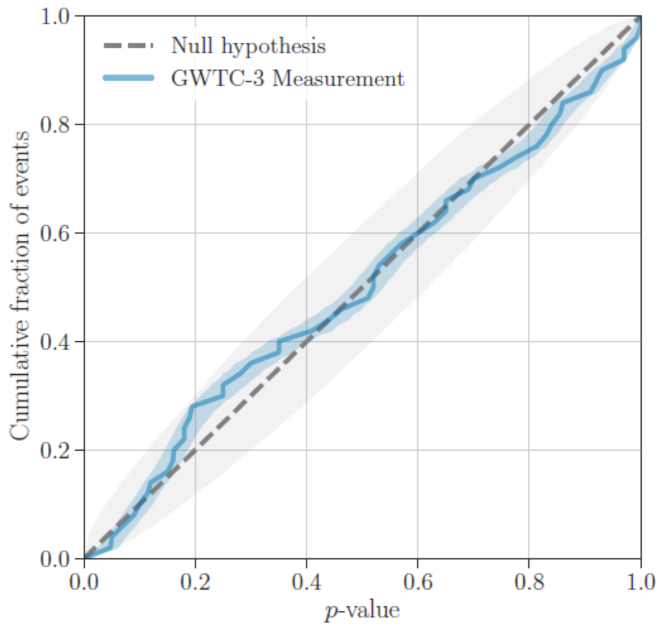
Important for model selection

Combining multiple observations

- We want to combine information from multiple events in order to
  - Infer the properties of the underlying source population
  - Test for deviations from general relativity
  - Infer the value of the Hubble constant
  - …
- Usually done an a subset of events, e.g. those with
  - Very low false-alarm rate
  - High SNR
  - High SNR in the ringdown
  - An electromagnetic counterpart, or good sky localization
  - …

With a pinch of frequentist analysis

- Study empirical distribution of some detection statistic for a frequentist null test of the hypothesis that GR is a good description of the data
  - e.g. residuals test: coherent network SNR after subtraction of best-fit GR waveform
- Compare detection statistic against empirical background distribution for each event
  - SNR computed on 200 randomly selected time segments around event time
  - p-value of residual SNR for each individual event
    - probability of obtaining a higher residual SNR from background
- Yields distribution of p-values
  - Under null hypothesis, p-values expected to be uniformly distributed in [0, 1]
- Comparison with expectation represented through probability–probability (PP) plot
  - Fraction of events with p-values ≤ given number
  - PP plot should be diagonal



arXiv:2112.06861

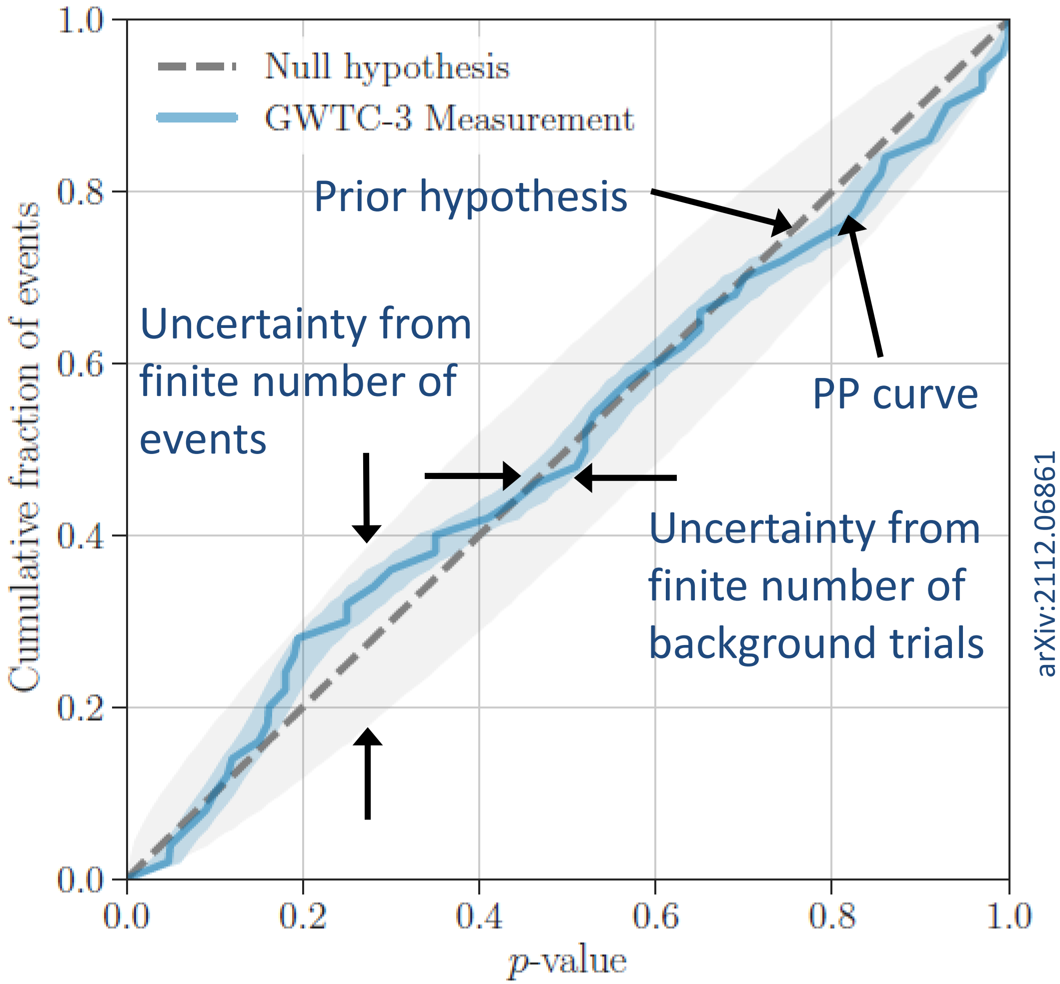
PP plot (cont.)

- N background trials around an event
- n give SNR higher than event
- Estimated p-value $\hat{p} = n/N$
- True p-value p
- Likelihood of $\hat{p}$ is binomial function

$$\mathcal{L}(\hat{p}) = \binom{N}{n} p^n (1-p)^{N-n}$$

- Posterior distribution of p

$$\mathrm{P}(p|N,n) = \mathrm{Beta}(n+1, N-n+1)$$



Hierarchical Bayesian Inference

- Use set of events to compare GR to beyond-GR model with extra parameter λ (GR: $\lambda = 0$)
  - e.g. parametrized post-Einstein framework
- Assume value of λ is the same for all events
  - Reasonable assumption in some cases (e.g. dispersion from massive graviton), too restrictive in most

$$p(\lambda|\boldsymbol{d}) \propto p(\lambda) \prod_i p(\boldsymbol{d}_i|\lambda)$$

- Assume value of λ is uncorrelated across events

$$\mathcal{B}_{\mathrm{GR/beyond-GR}} = \prod_i \mathcal{B}_{\mathrm{GR/beyond-GR,i}}$$

- General case: assume λ is drawn from an unknown distribution

$$p(\lambda|\mu,\sigma) \sim \mathcal{N}(\mu,\sigma) \qquad p(\lambda|\boldsymbol{d}) = \int p(\mu,\sigma|\boldsymbol{d}) p(\lambda|\mu,\sigma) d\mu d\sigma$$

  - GR: $\mu = 0$ & $\sigma = 0$ Previous cases: $\mu \neq 0$ & $\sigma = 0$ or $\sigma = \infty$

Inferring an astrophysical population

- Use set of events to infer e.g. mass distribution of sources
- Also based on hierarchical Bayesian inference, but selection effects need to be taken into account
  - Observed population has Malmquist bias
    - Loudest sources more likely to be detected

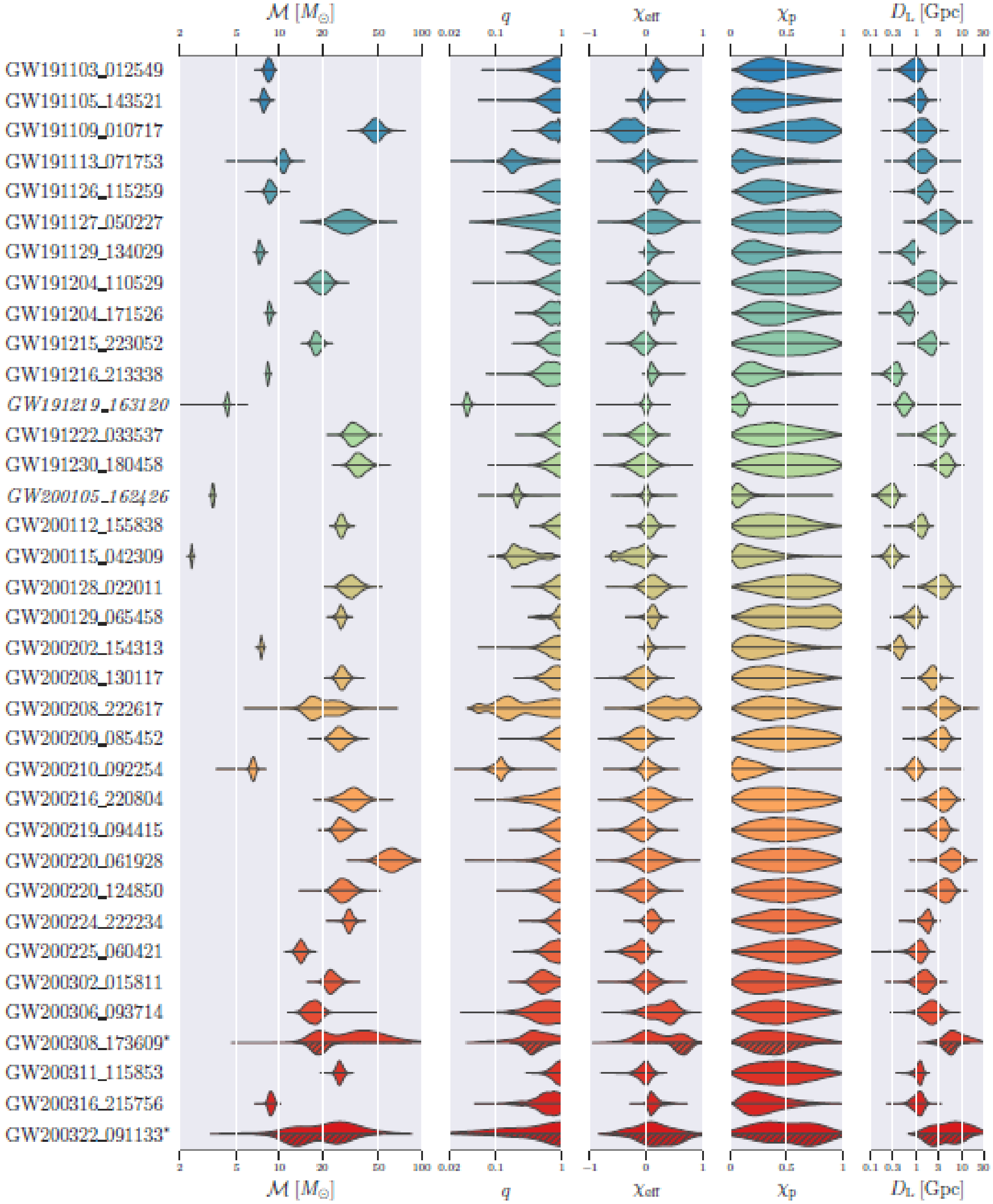
Likelihood of i^{th} event data under parameters θ

$$\mathcal{L}(\{d\}|\Lambda) \propto \prod_{i=1}^{N_{\rm det}} \frac{\int \mathcal{L}(d_i|\theta)\pi(\theta|\Lambda)d\theta}{\xi(\Lambda)}$$

Distribution of event parameters θ for population with parameters Λ

Population model parameters

Fraction of detectable events for population with parameters Λ



$\mathcal{M}$ [$M_\odot$]
q
$\chi_{\rm eff}$
$\chi_{\rm p}$
$D_{\rm L}$ [Gpc]
GW191103_012549
GW191105_143521
GW191109_010717
GW191113_071753
GW191126_115259
GW191127_050227
GW191129_134029
GW191204_110529
GW191204_171526
GW191215_223052
GW191216_213338
GW191219_163120
GW191222_033537
GW191230_180458
GW200105_162426
GW200112_155838
GW200115_042309
GW200128_022011
GW200129_065458
GW200202_154313
GW200208_130117
GW200208_222617
GW200209_085452
GW200210_092254
GW200216_220804
GW200219_094415
GW200220_061928
GW200220_124850
GW200224_222234
GW200225_060421
GW200302_015811
GW200306_093714
GW200308_173609*
GW200311_115853
GW200316_215756
GW200322_091133*