Coalescing Compact Binaries: From first discoveries to populations

John Veitch
University of Glasgow

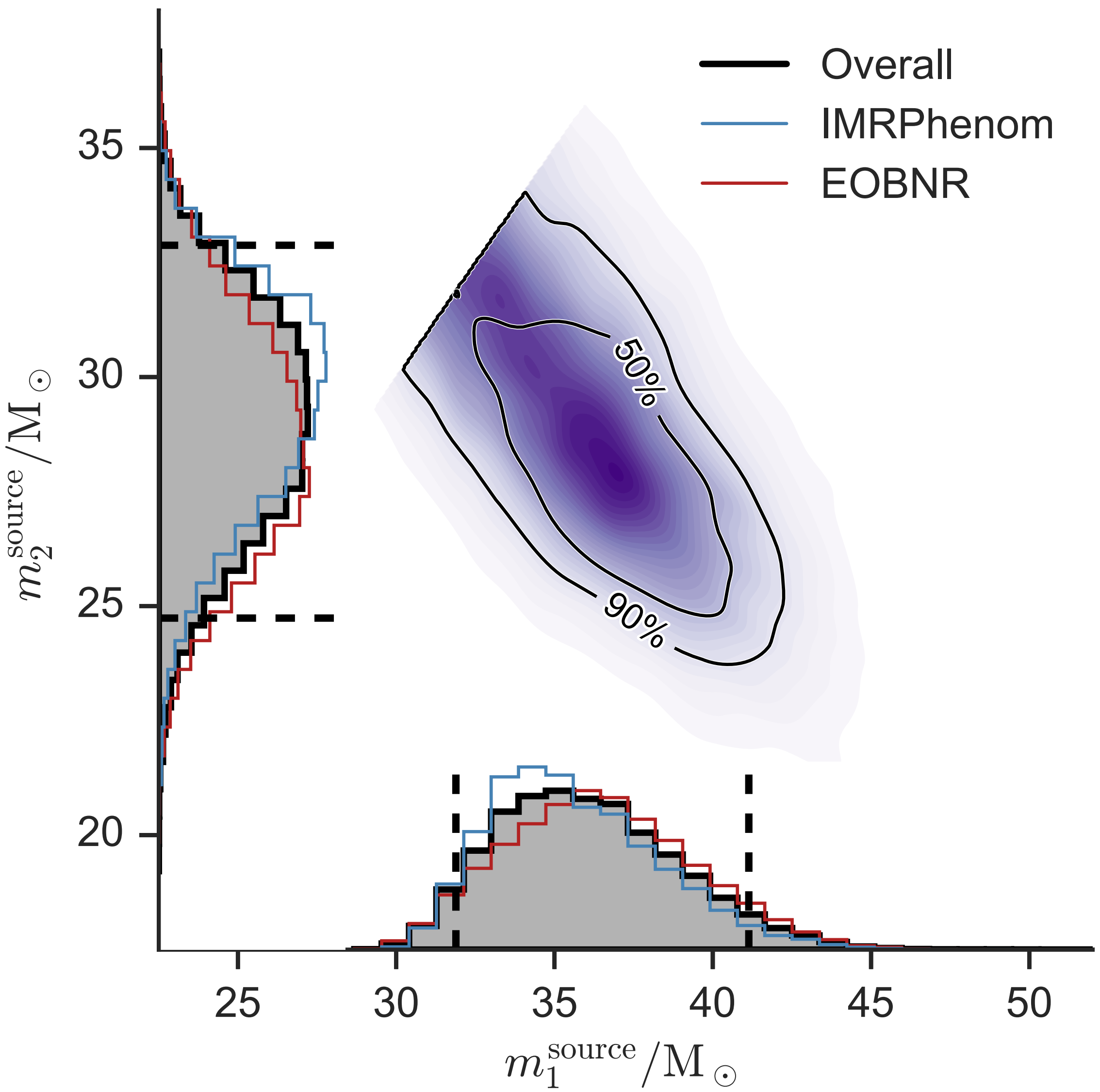
Pisa Seminar
2018-08-05

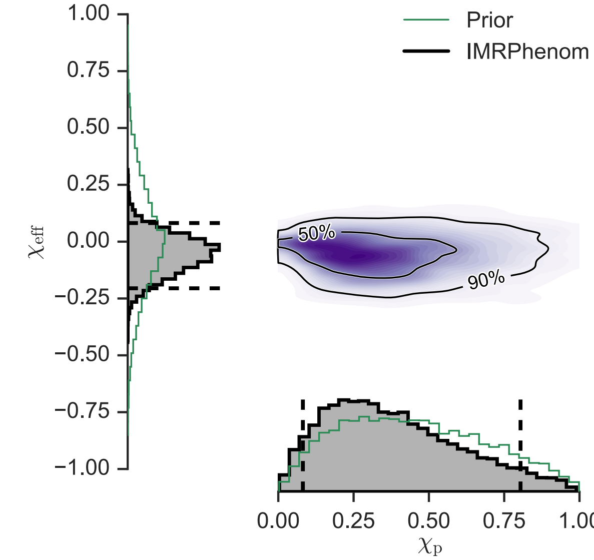
LSC VIRGO

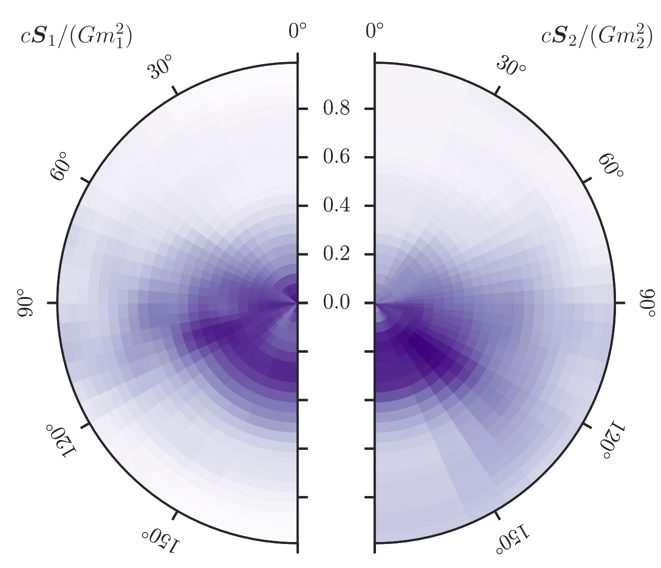
University of Glasgow

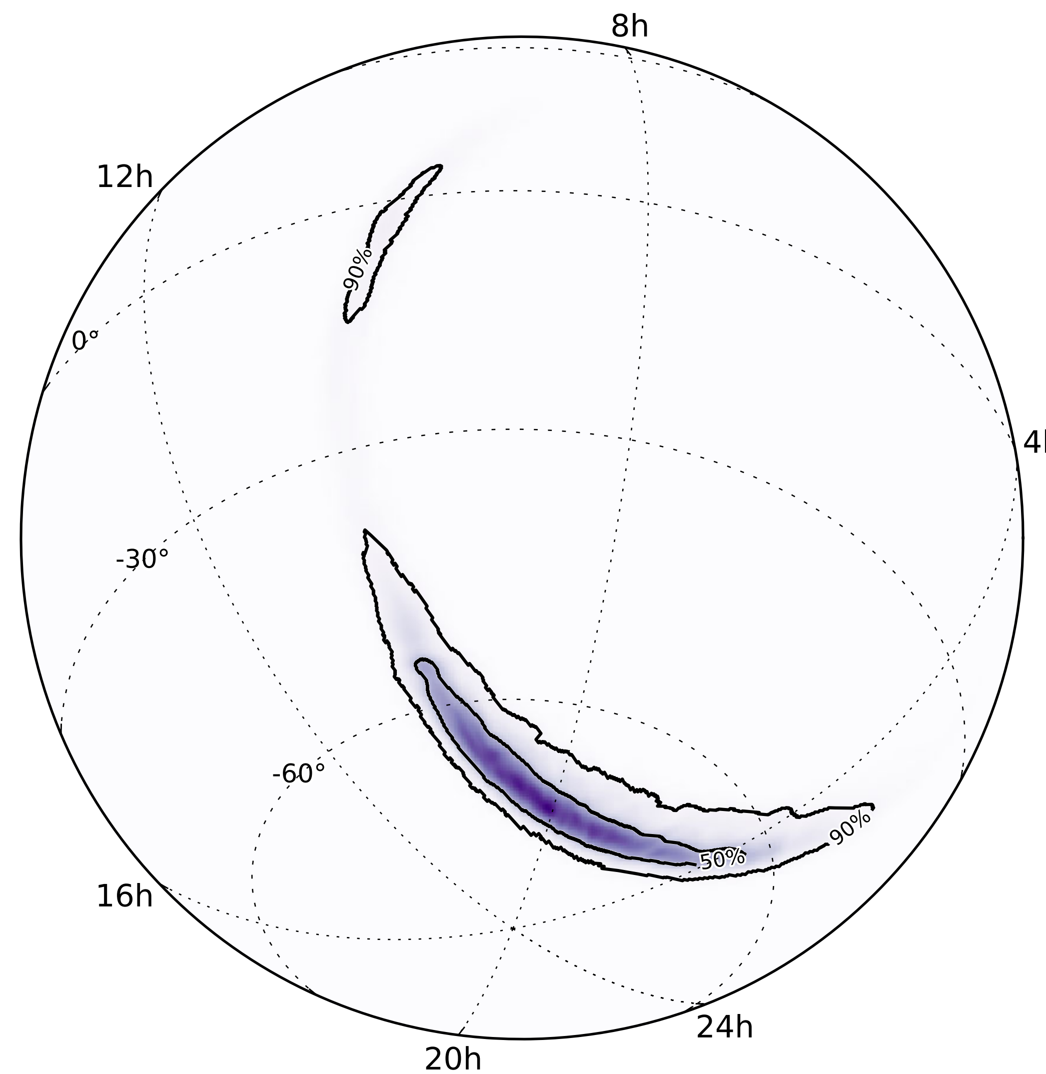
Contents

- Extracting source parameters from GW data
- Results from O1 and O2 to date
- Population analyses


Overall
IMRPhenom
EOBNR
50%
90%
$m_2^{\mathrm{source}}/\mathrm{M}_\odot$
$m_1^{\mathrm{source}}/\mathrm{M}_\odot$

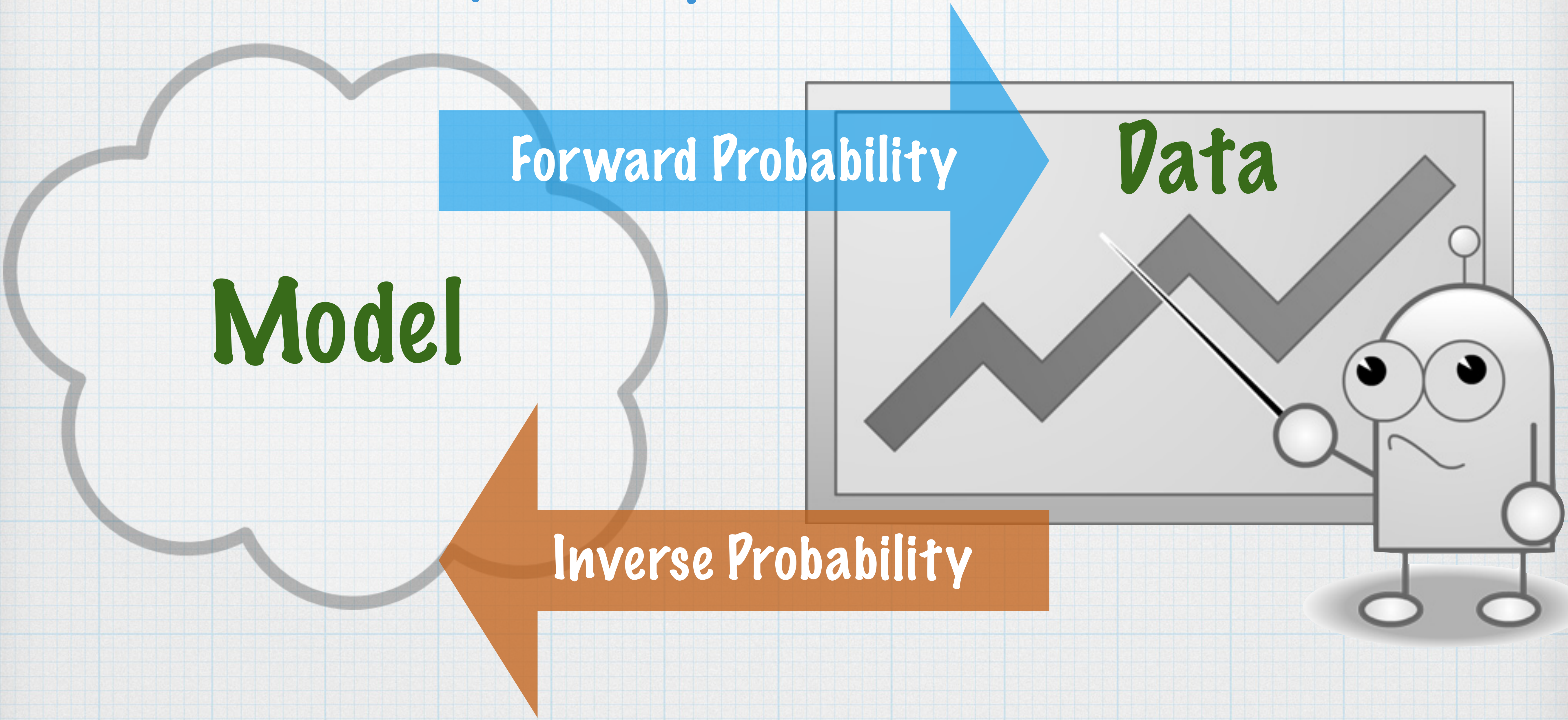

Prior
IMRPhenom
χ_{eff}
χ_{p}
50%
90%


$cS_1/(Gm_1^2)$
$cS_2/(Gm_2^2)$


8h
12h
4h
16h
20h
24h
0°
-30°
-60°
90%
50%

Forward and Inverse Probability

Forward Probability - make predictions about data from a model



Inverse Probability - Learn about a model from data

Bayes' Theorem

- What we need is a method to update our beliefs given new information

- From product rule comes Bayes's theorem

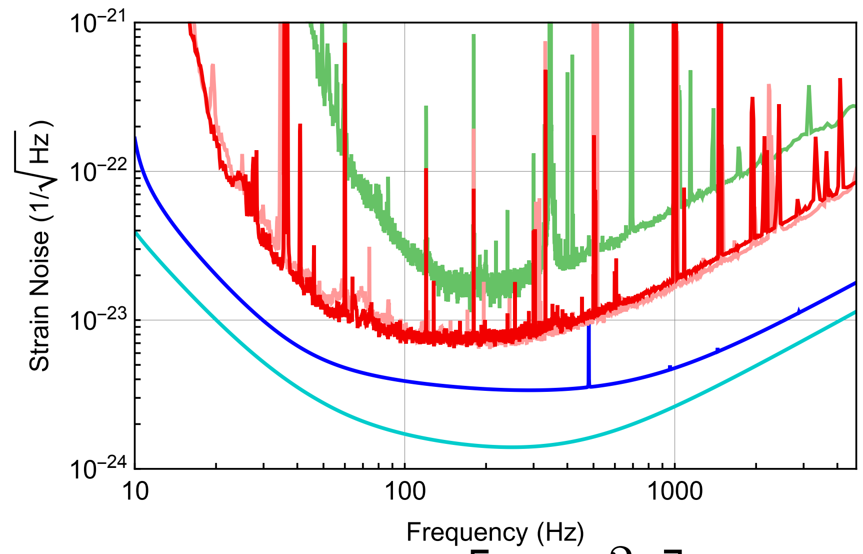
$$P(\text{theory}|\text{new data}) = \frac{P(\text{theory})P(\text{new data}|\text{theory})}{P(\text{new data})}$$

- Together with the axioms gives us an algorithmic method for updating our state of knowledge given some new information

Parameter Estimation

- We express knowledge about a parameter's value through a *probability density function*.
    - p(*x*|I) = probability density of the parameter *x*
- Normalised: $$\int p(x|I)dx = 1$$
- $$P(A < x < B|I) = \int_A^B p(x|I)dx$$

GW noise model



- Simplest noise model makes a few assumptions:
  - zero mean: $\langle n_i \rangle = 0$
  - known variance: $\langle n_i^2 \rangle = S_h(f_i)/\Delta f = \sigma_i^2$
- Maximum entropy distribution is Gaussian, i.e. $p(n_i|\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{n^2}{2\sigma^2}\right]$
  - Assuming stationarity implies independence in each frequency bin:

$$p(\{n_i\}|\{\sigma_i\}) = \prod_i p(n_i|\sigma_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{n^2}{2\sigma^2}\right]$$

  - And independence in each detector:

$$p(\vec{n_H}, \vec{n_L}, \vec{n_V}|\vec{\sigma_H}, \vec{\sigma_L}, \vec{\sigma_V}) = p(\vec{n_H}|\vec{\sigma_H})p(\vec{n_L}|\vec{\sigma_L})p(\vec{n_V}|\vec{\sigma_V})$$

  - Terminology: the “likelihood” of the noise given σ

Likelihood function

- For additive noise, $d_i = h_i + n_i$, the mean of the data distribution becomes the prediction of the signal model for given parameters θ: $\langle d_i \rangle = h(f_i, \theta)$, whereas variance remains the same.

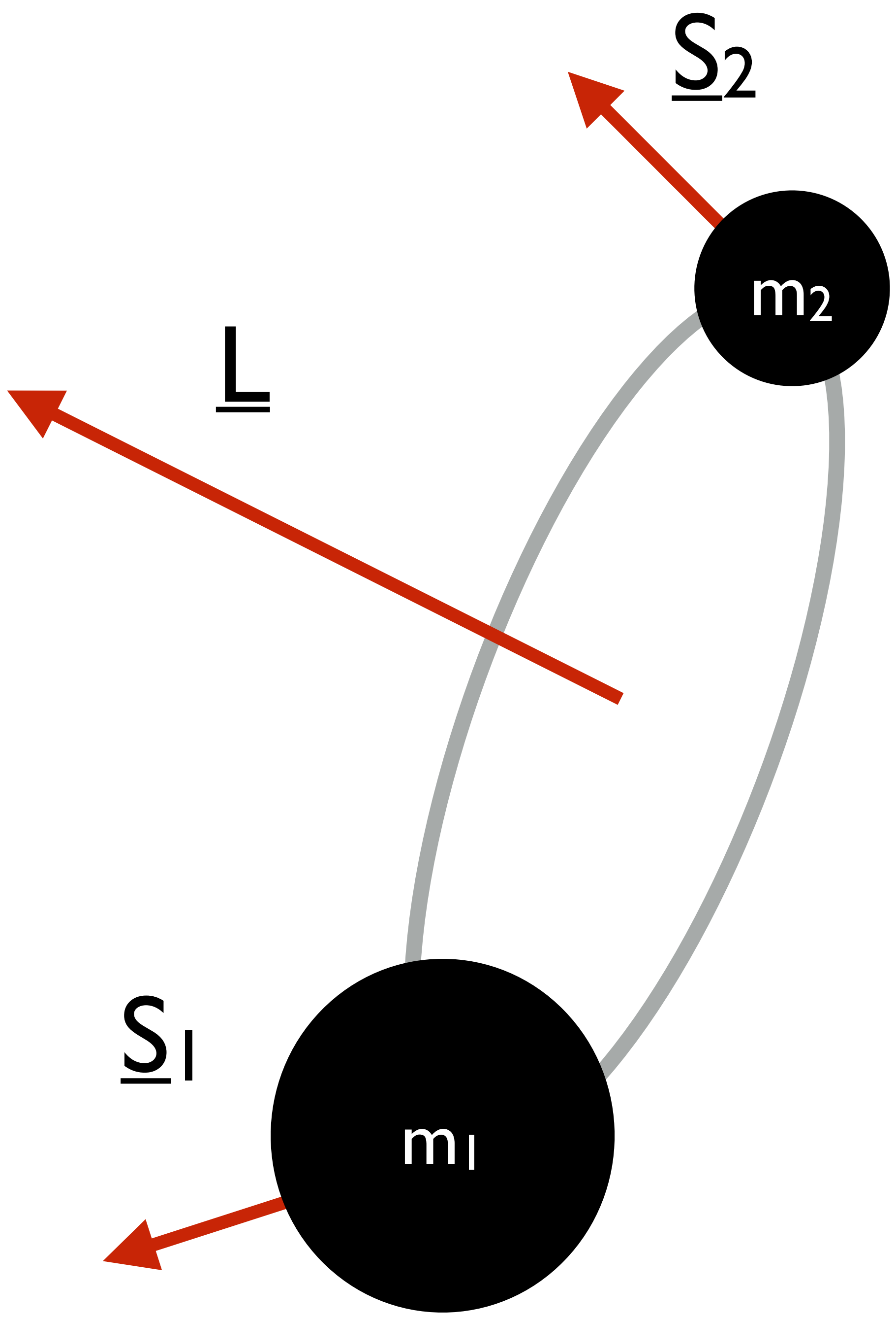
- If we then observe a specific set of data, we can infer the parameters of the signal by calculating the posterior probability distribution (PDF):

$$p(\vec{\theta}|\vec{d}, \vec{\sigma}, H_S) = \frac{p(\vec{d}|\vec{\theta}, \vec{\sigma}, H_S)p(\vec{\theta}|\vec{\sigma}, H_S)}{p(\vec{d}|\vec{\sigma}, H_S)}$$

- where $p(\vec{d}|\vec{\sigma}, H_S) = \int d^N \vec{\theta} p(\vec{d}|\vec{\theta}, \sigma, H_S)p(\vec{\theta}|\vec{\sigma}, H_S)$
is the *evidence* or marginal likelihood of the model.

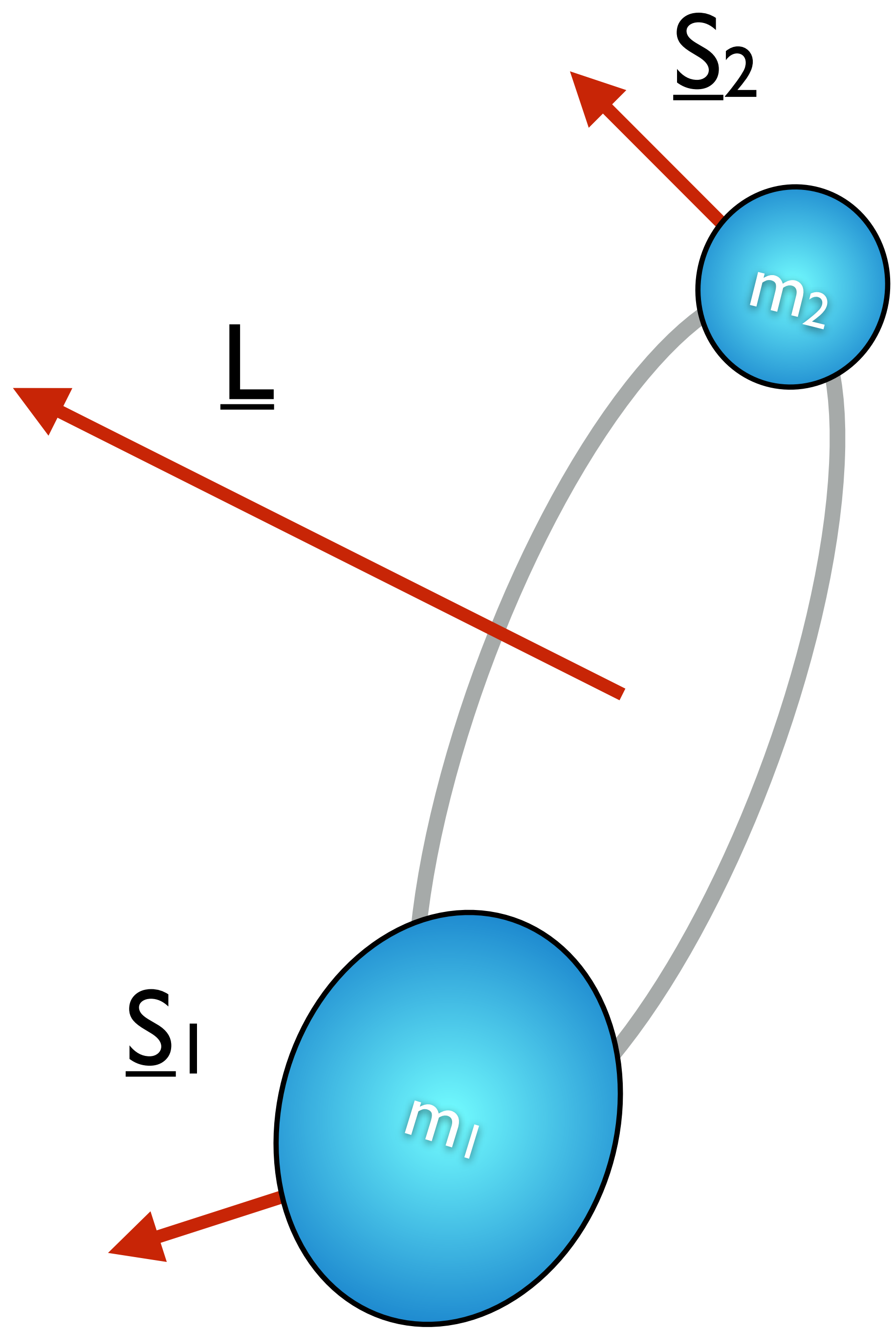
What is there to measure?

- Intrinsic Parameters
    - masses
    - spins
- Extrinsic Parameters
    - Inclination
    - Orientation
    - Polarisation
    - Sky position
    - luminosity distance
    - time



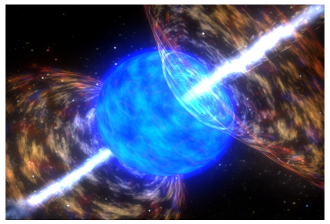
What is there to measure?

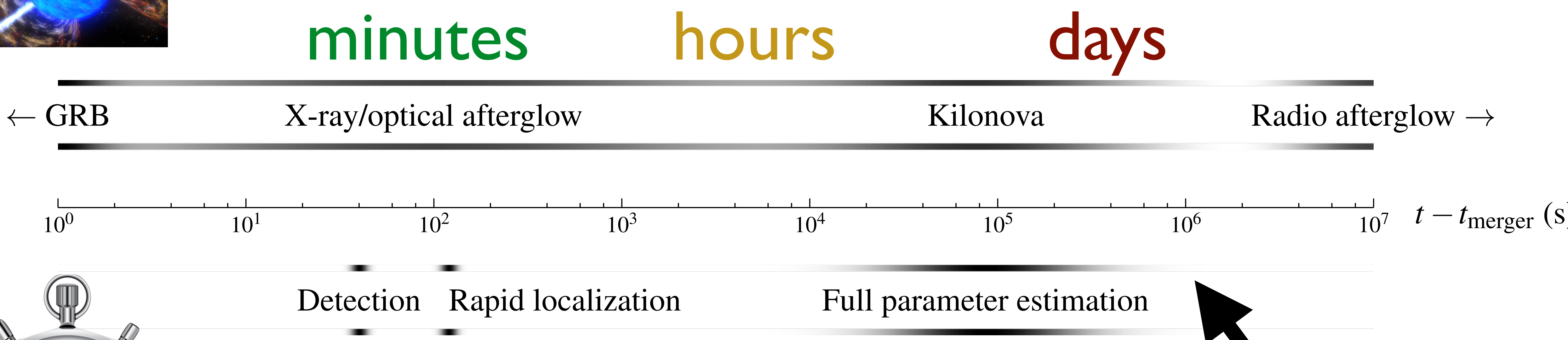
- **Subtler effects**
- NS Equation of state
  - tidal deformation
- Deviations from GR
- eccentricity



LSC VIRGO University of Glasgow

Speed vs complexity

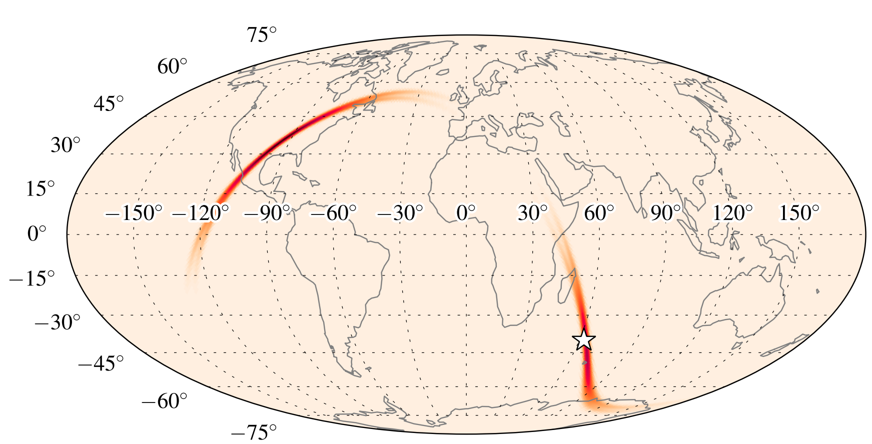




rapid localization

fast, non-spinning

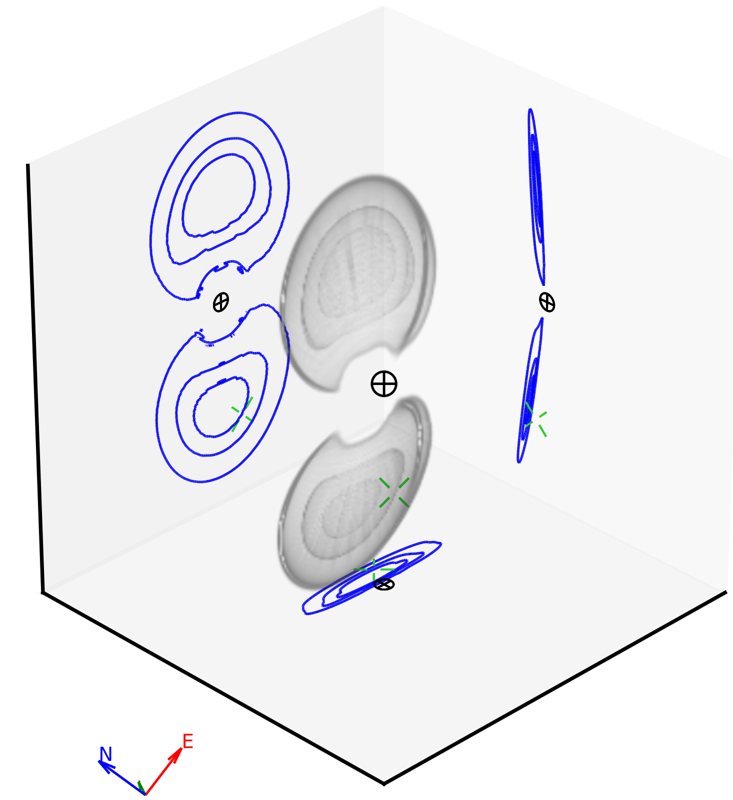
slow,
fully spinning



[Singer+ 2014 ApJ 795 105]
[Berry+ arXiv:1411.6934]

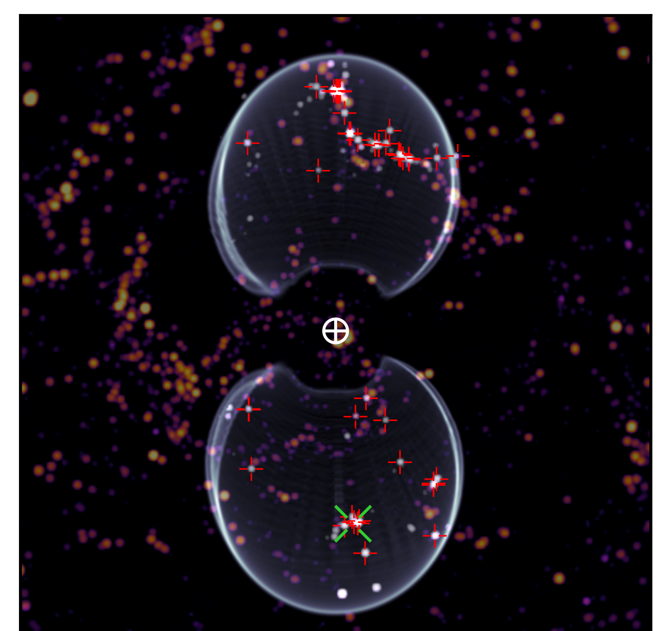
Rapid Localisation

- BAYESTAR `[Singer+ ApJ 795 (2014), Singer+ ApJ. L15 829 2016]` algorithm fixes masses to best fit template from search - reduces parameter space dimension to extrinsic



$$p(\alpha, \delta, \psi, \phi_0, \iota, d_L | m_1, m_2, \vec{s}_1, \vec{s}_2)$$

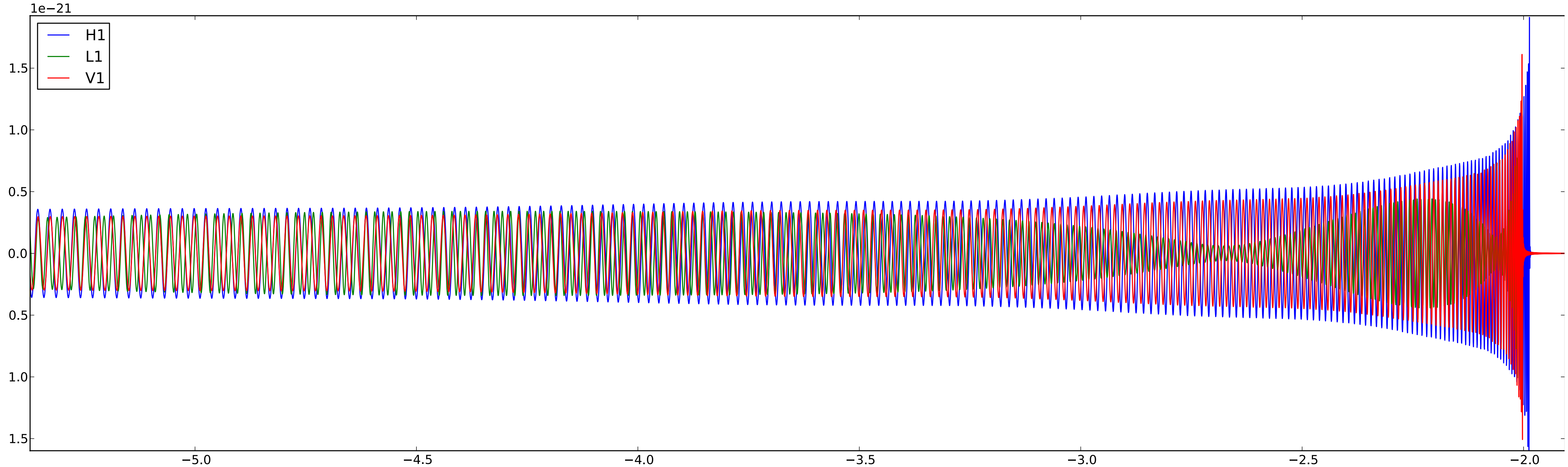
- Marginalisation over inclination, polarisation, phase to localise in area and volume using numerical integration and lookup tables



$$p(\alpha, \delta, d_L | d, m_1, m_2, \vec{s}_1, \vec{s}_2) = \int p(\alpha, \delta, \psi, \phi_0, \iota, d_L | d, m_1, m_2, \vec{s}_1, \vec{s}_2) d\psi d\phi_0 d\iota$$

- Rapid identification of EM counterpart!

Full PE



- For full parameter estimation, need to explore 9 (non-spinning), 11 (aligned spin), 15 (precessing spin) or more (tidal, non-GR) parameters
- A typical signal can contain ~30-40 bits of information
  - Test 2^{30} points in parameter space to find peak of PDF!
- Non-linear signal model makes general linear solution impossible
- Use stochastic sampling techniques to draw *samples* from the posterior distribution function

LSC VIRGO

University of Glasgow

LALInference

- LALInference library developed by LIGO-Virgo collaboration [Veitch+ 2015] for stochastic sampling
  - Free software https://git.ligo.org/lscsoft/lalsuite
  - Makes use of state of the art waveforms in LALSimulation
  - Reviewed and robust waveforms, likelihoods, priors, samplers.
- Uses MCMC and Nested Sampling for parameter estimation & model selection

Markov Chain Monte Carlo

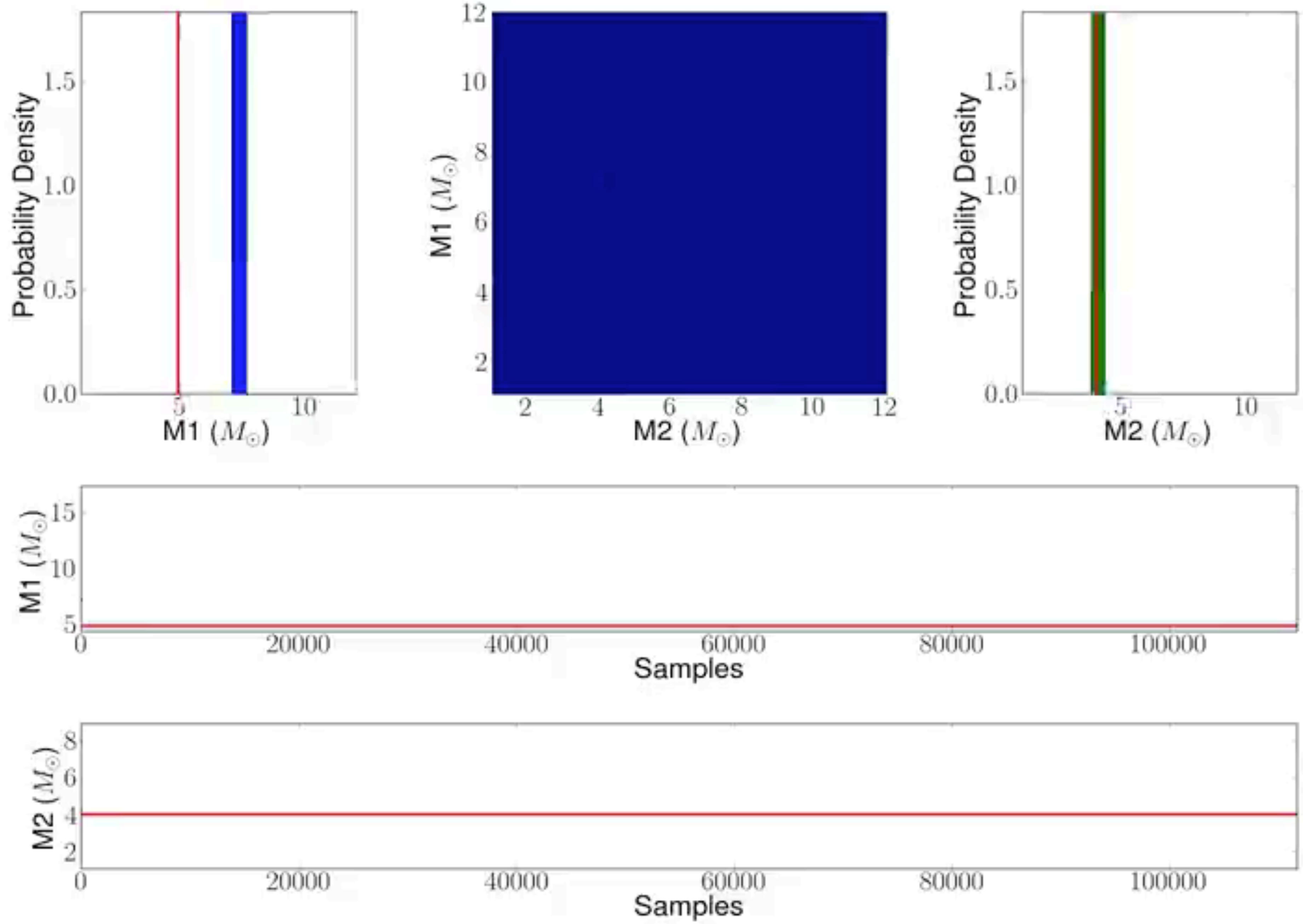
- MCMC is a technique to draw samples from a target distribution p(x|d) (in our case the posterior distribution) sequentially
- Given a starting point x_0, a new sample is drawn from a proposal distribution $q(x'|x_0)$ which can depend only on the last point (forgets its history)
- It is then accepted with acceptance probability

$$\alpha(x'|x_0) = \min\left[1, \frac{p(x'|d)q(x_0|x')}{p(x_0|d)q(x'|x_0)}\right]$$

- or otherwise the previous value x_0 is repeated
- This is the Metropolis-Hastings algorithm

MCMC

Random walk around parameter space with Metropolis-Hastings algorithm targeting the 2D mass posterior



UNIVERSITY OF BIRMINGHAM

VIRGO LSC

Extensions to MCMC algorithm

- Enhanced algorithms can dramatically improve the efficiency by better exploring parameter space

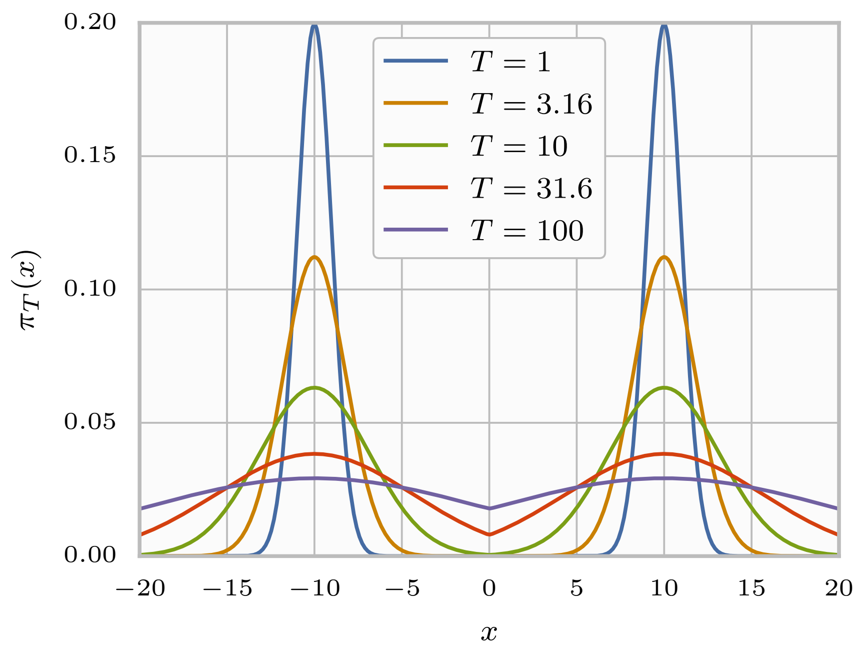
- A few examples used by LALInference…

Parallel Tempering

- If posterior is sharply peaked can be difficult to find high probability region

$$\pi_T(\vec{\theta}) \propto L(\vec{\theta})^{1/T} p(\vec{\theta})$$

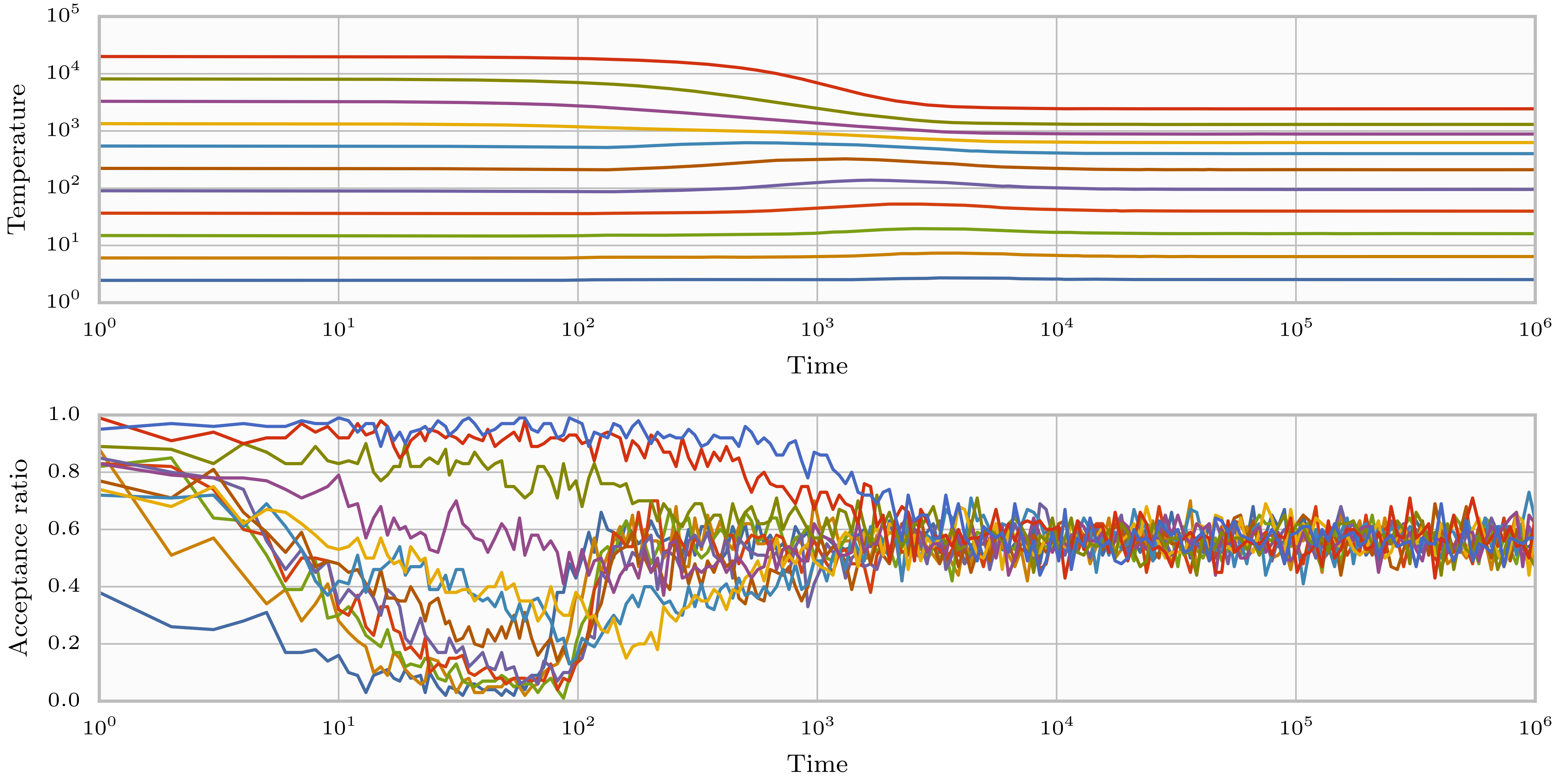
- *Anneal* the likelihood function



- Run multiple chains at different temperatures and allow jump proposals between them

$$A_{i,j} = \min\left\{ \left(\frac{L(\vec{\theta}_i)}{L(\vec{\theta}_j)} \right)^{\beta_j - \beta_i}, 1 \right\}$$

Adaptive Parallel Tempering

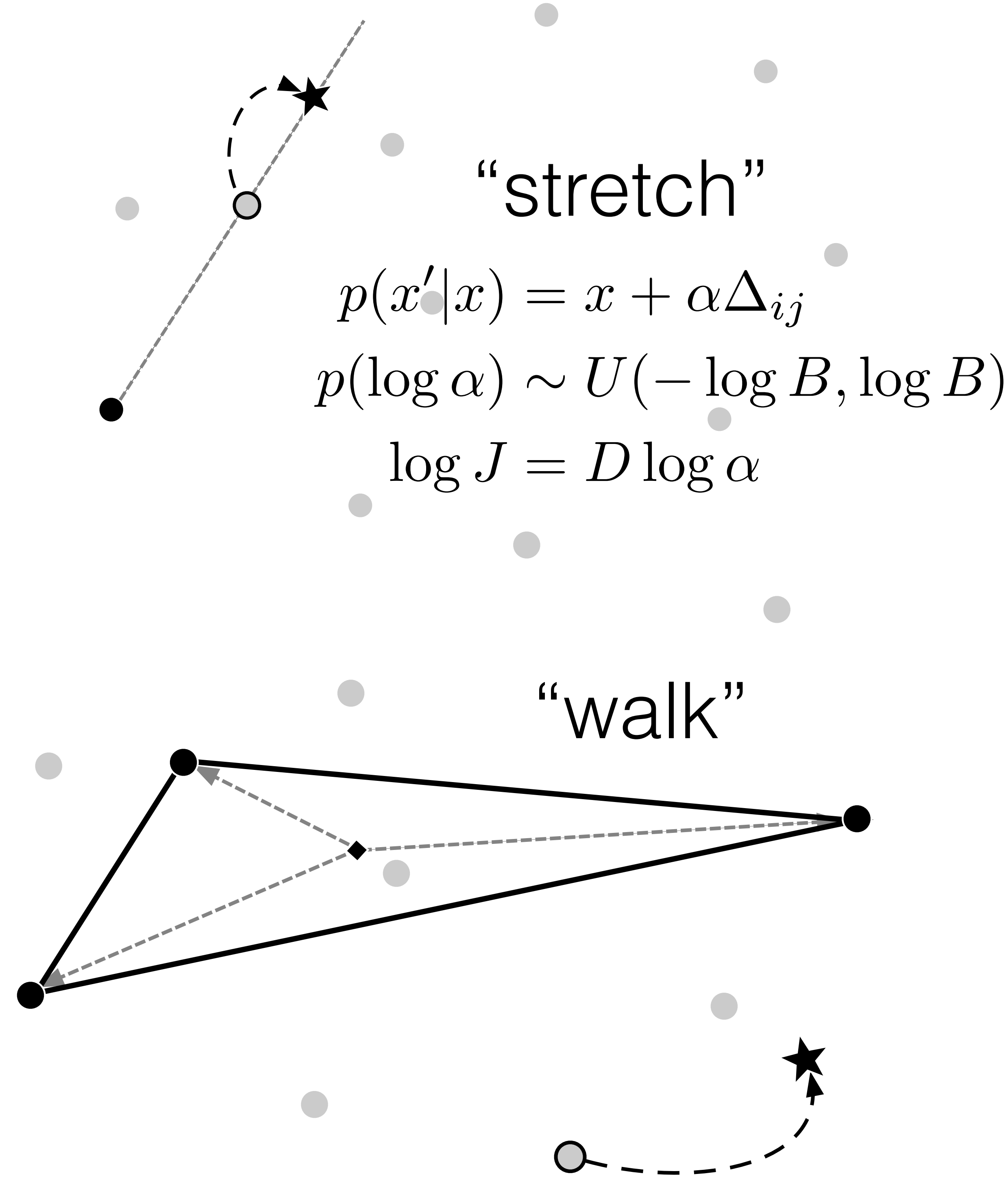


Vousden+ 1501.05823

University of Glasgow

Ensemble Samplers

- Goodman & Weare 2009 - also Emcee
- Evolve a collection of samples instead of single point
- At each iteration, collection is a sample from the target distribution
- This is also useful for Nested Sampling method!



LSC VIRGO University of Glasgow

Model Selection

* To test different hypotheses we can compute the ratio of their probabilities P(A)/P(B), known as the "odds ratio"

* Given some experimental data we can update our relative belief in two models

$$\frac{P(A|d)}{P(B|d)} = \frac{P(A)}{P(B)} \frac{P(d|A)}{P(d|B)}$$

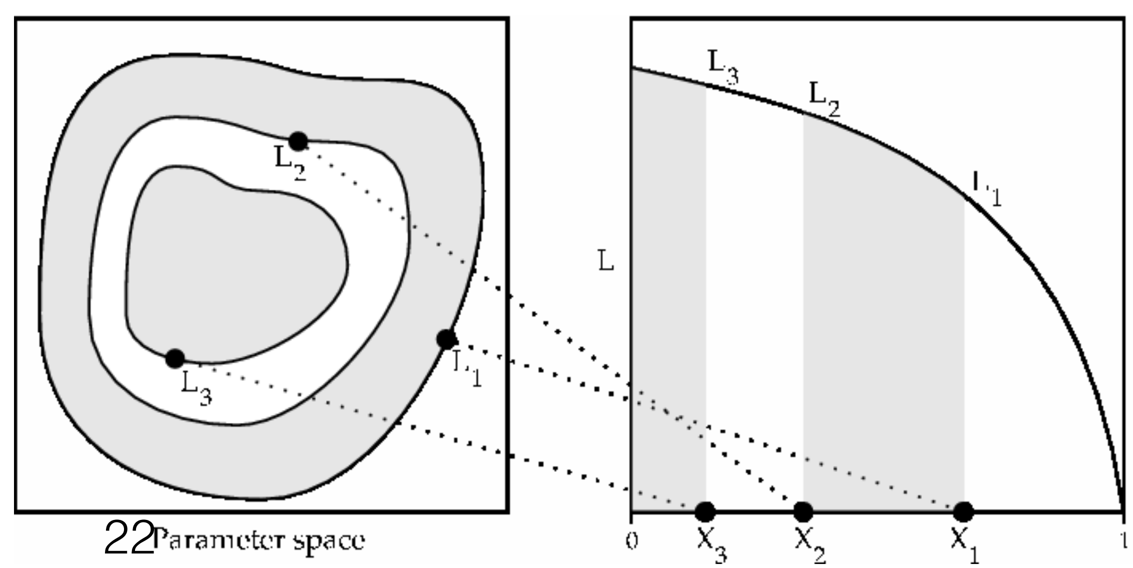
* The evidence ratio is also known as the "Bayes Factor"

$$\frac{P(d|A)}{P(d|B)} = \frac{\int_X p(d,x|A)dx}{\int_Y p(d,y|B)dy}$$

* Allows comparison of models of different parameter spaces X and Y

Nested Sampling

- We use Nested Sampling to evaluate the integral probabilistically.
- From product rule, get $p(d|\vec{\theta},H,I)p(\vec{\theta}|H,I)d\vec{\theta} = P(d|H,I)p(\vec{\theta}|d,H,I)d\vec{\theta}$
- Write element of prior as $dX = p(\vec{\theta}|H,I)d\vec{\theta}$ and note
- Evidence is then given $Z = P(d|H,I) = \int p(d|\vec{\theta},H,I)dX = \int L dX$
- Define $X(\lambda) = \int_{L(\vec{\theta})>\lambda} dX$ as the prior probability mass covering the regions of parameter space with likelihood >λ. As λ increases, enclosed probability X shrinks from 1 to 0. Using inverse function evidence is
- Now to do 1-D integral....



Nested Sampling

- If we had a collection of m points from the prior, we could approximate the integral as Riemann sum $Z = \sum_{i=1}^{m} L_i w_i$ where $w_i = \Delta X$

- But how do we get from θ to X? Sample m points X_i randomly from the parameter space, and calculate their likelihoods L_i, then sort them by L. Lowest L corresponds to highest X.

- At each iteration, replace lowest point (L_{min} , X_{max}) by another sampled from within the limited prior $X(L_{min})$. At each iteration X therefore decreases geometrically,

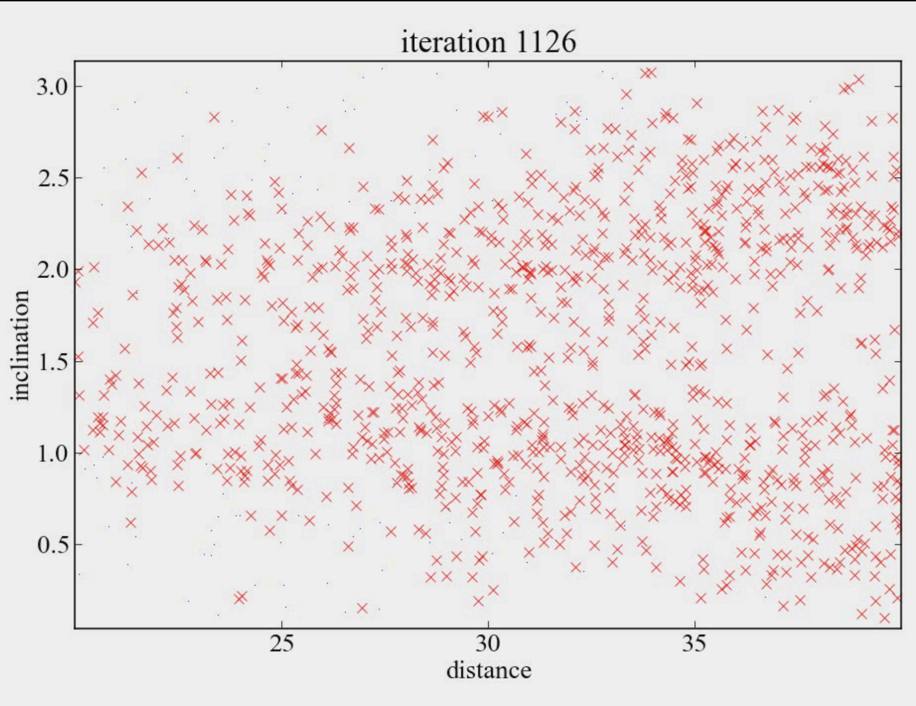
$$X_i = t_i X_{i-1}, \qquad < \log t_i >= -m^{-1}$$

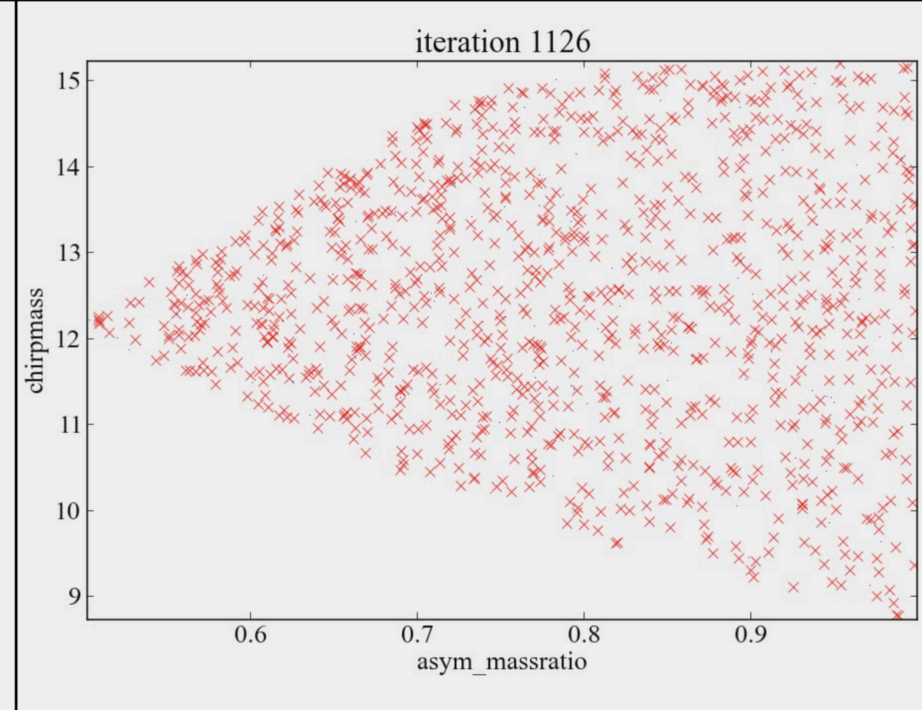
- To sample the limited prior, design a special MCMC routine - we can use the output samples to do parameter estimation too!

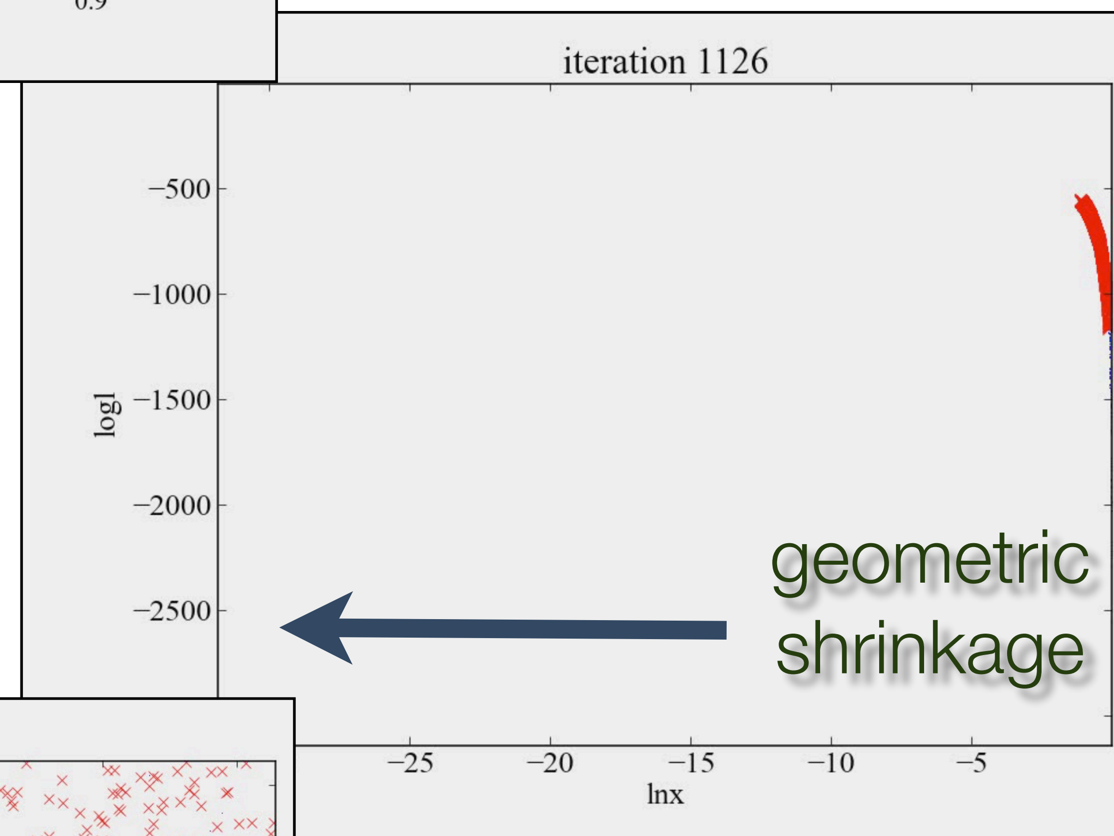
- Recalling that $X_0=1$, we can now evaluate the above sum by iterating the process, and find the evidence!

Nested Sampling

[Veitch & Vecchio PRD 81 (2010)]



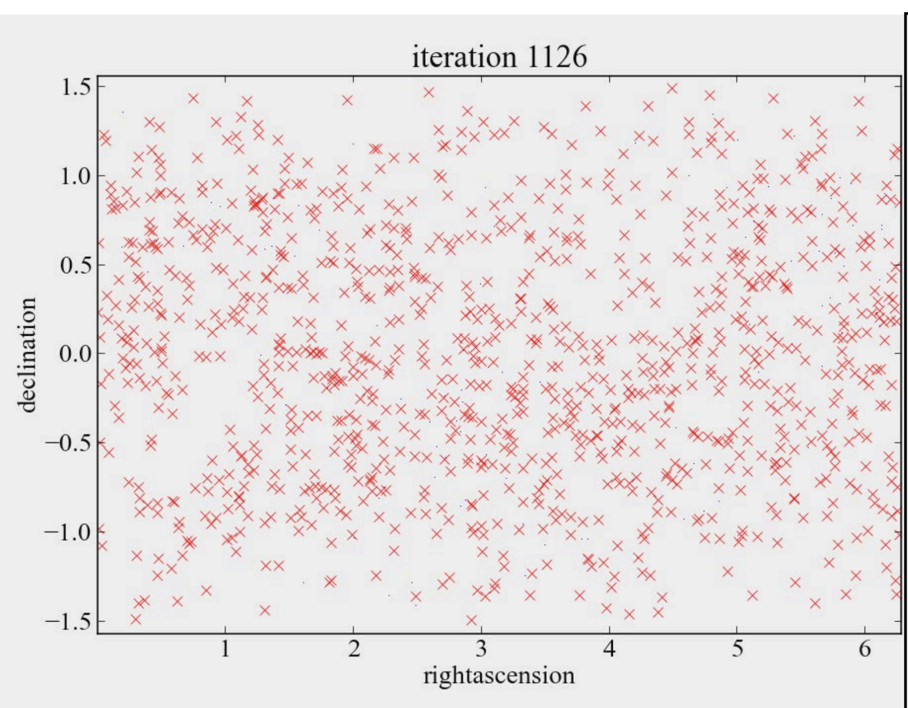


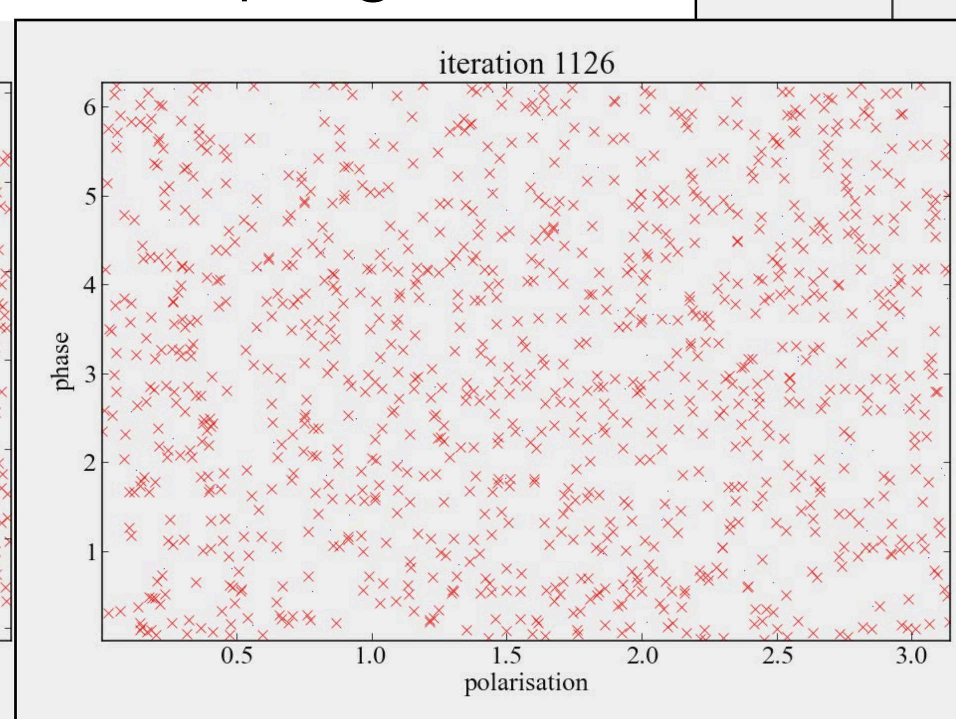


Live points (located inside a contour of constant likelihood)

Used points

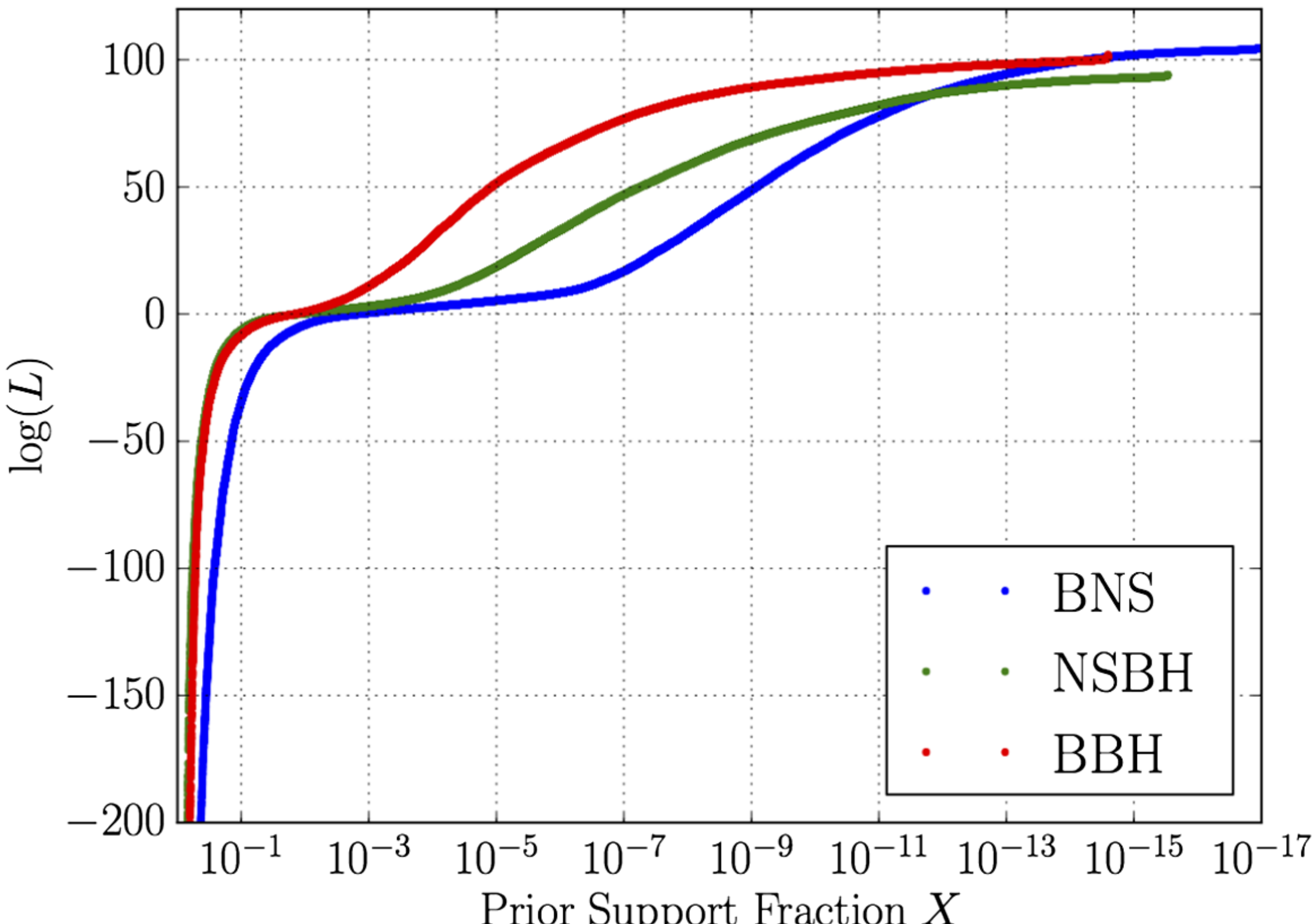
As the contours shrink, replacement samples are drawn from inside the contour using MCMC or rejection sampling



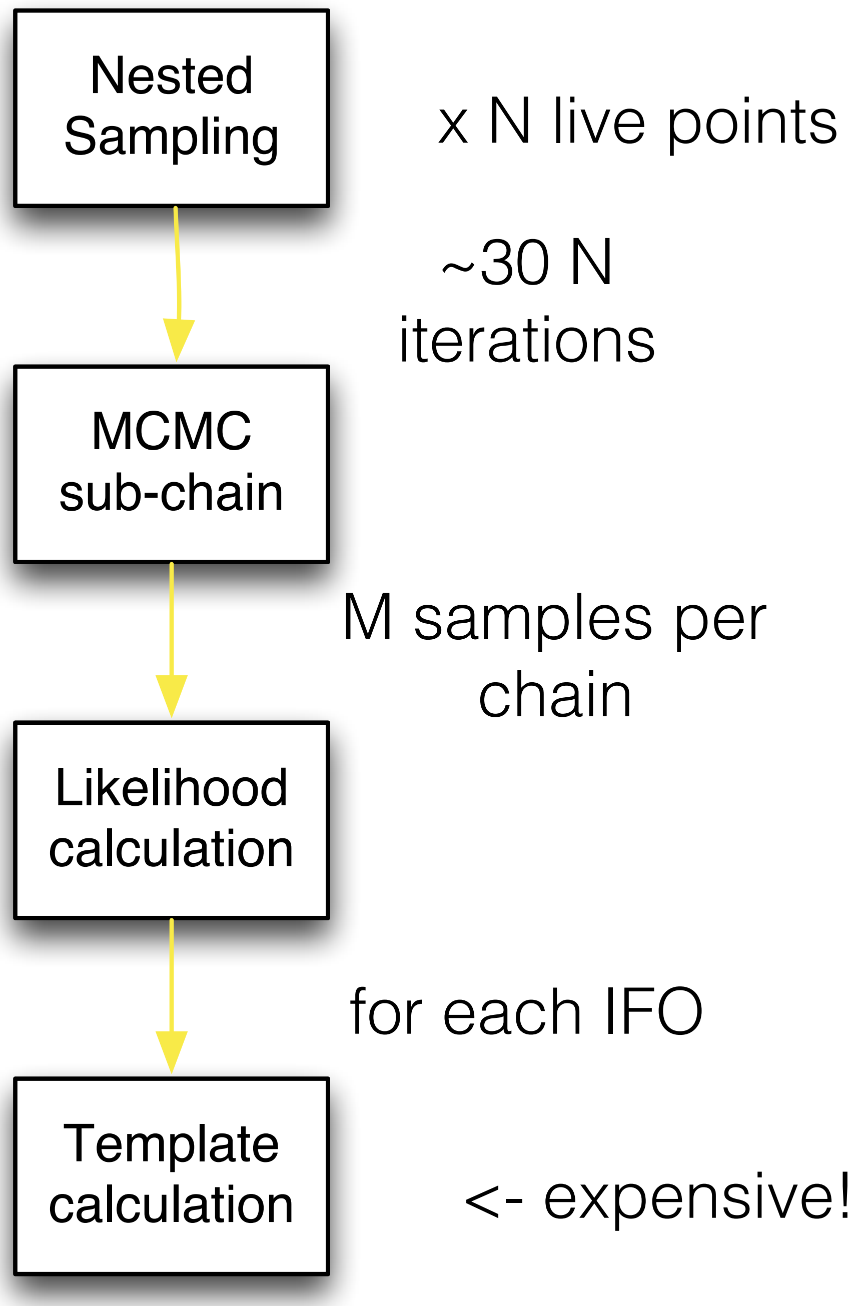


Example CBCs

BBH easier to sample than BNS!



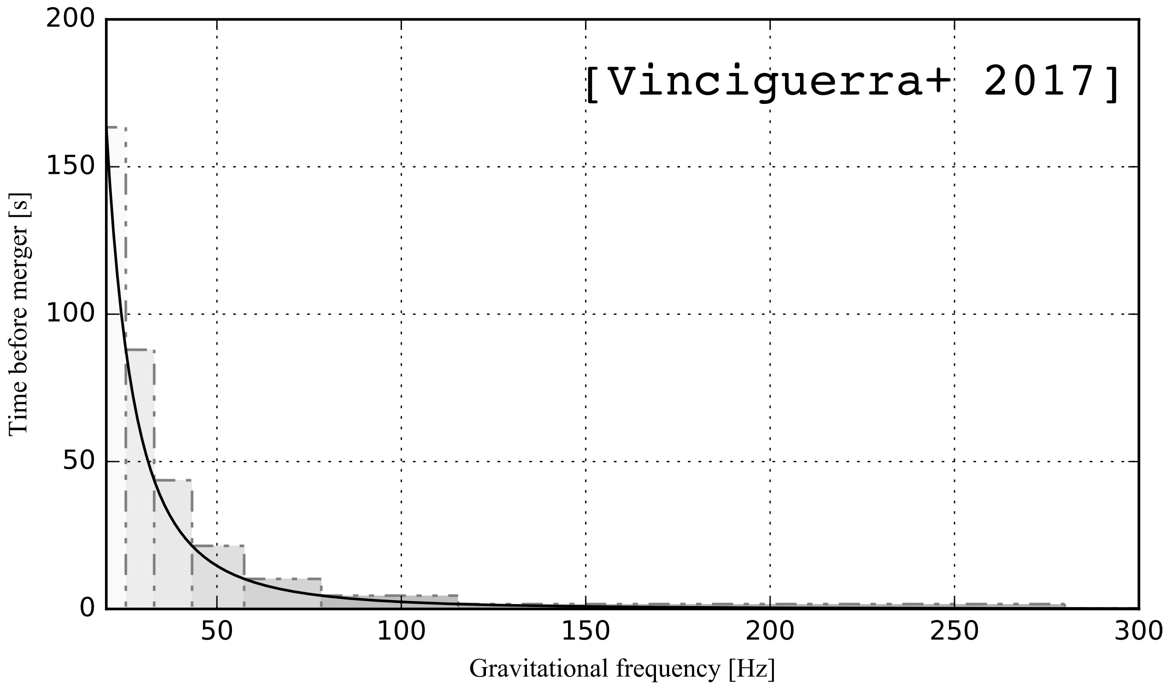
Computational Cost



- Run-time is dominated by
  - Template calculation (60-99% run-time)
    - including FFT for time-domain signals (rarely used in practice)
    - scales as length of template x sampling frequency
  - Overlap calculation (remainder of run-time)
    - Once per detector
    - Uses accelerated trigonometry approximation
- Reduce number of likelihoods needed
- Parallelise multiple runs across cores to reduce wall clock time
- Improve computational efficiency of vector ops (SIMD extensions)
- Use of custom hardware? GPUs, Xeon Phi, …

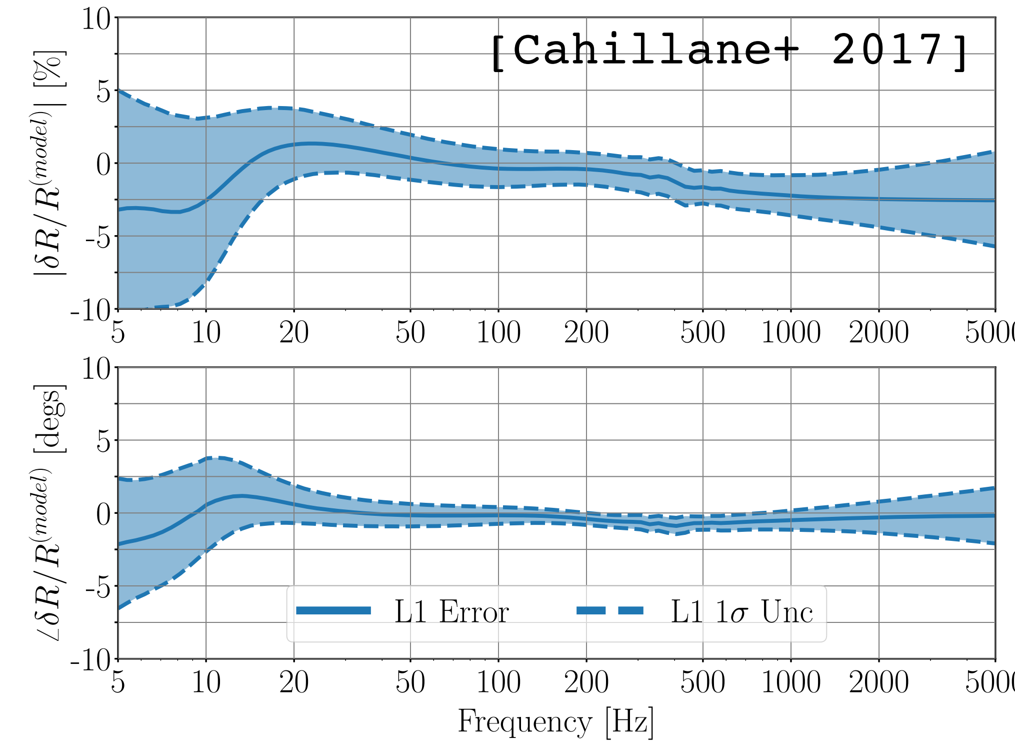
Waveform acceleration

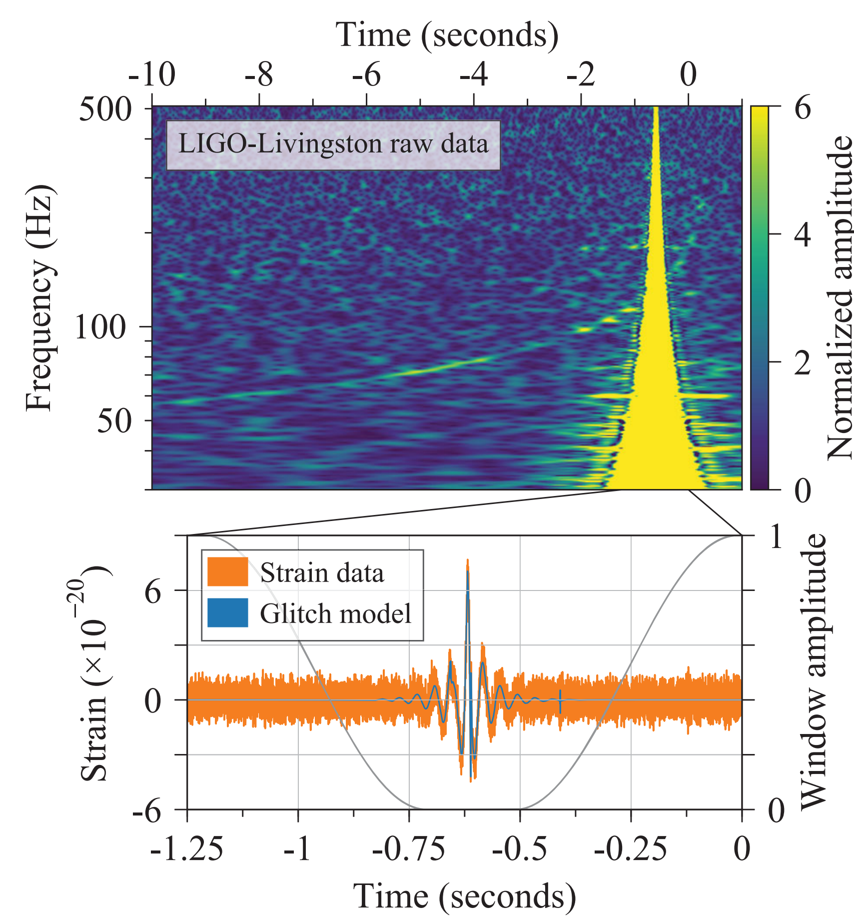
- Dominant cost is generation of ~10M waveforms while exploring parameter space
- Most sophisticated inspiral-merger-ringdown waveforms are slow to generate
  - Surrogate models to approximate full waveform `[Pürrer 2014]`
  - Multi-band analysis `[Vinciguerra+ 2017]`
- Use basis other than Fourier frequencies for likelihood integrals “Reduced Order Quadrature” `[Canizares+ 2014, Smith+ 2016]`



Reducing systematics

- Systematic errors caused by…
- detector calibration [Cahillane+ 2017]:
    - Include calibration uncertainty as parametrised model [Vitale+ 2012, LVC LVC PRL 116 2 (2016)]
- Imperfect templates [LVC 1611.07531]
    - Gaussian process modelling of waveform uncertainties [Moore+ 2014, 2015]
- Varying noise floor (PSD estimation)
    - On-source PSD estimation [Littenberg+ 2014]
- Glitches in data! [e.g. LVC, PRL 119 (2017)]
    - Fit glitch model along with GW [Cornish+ 2014]



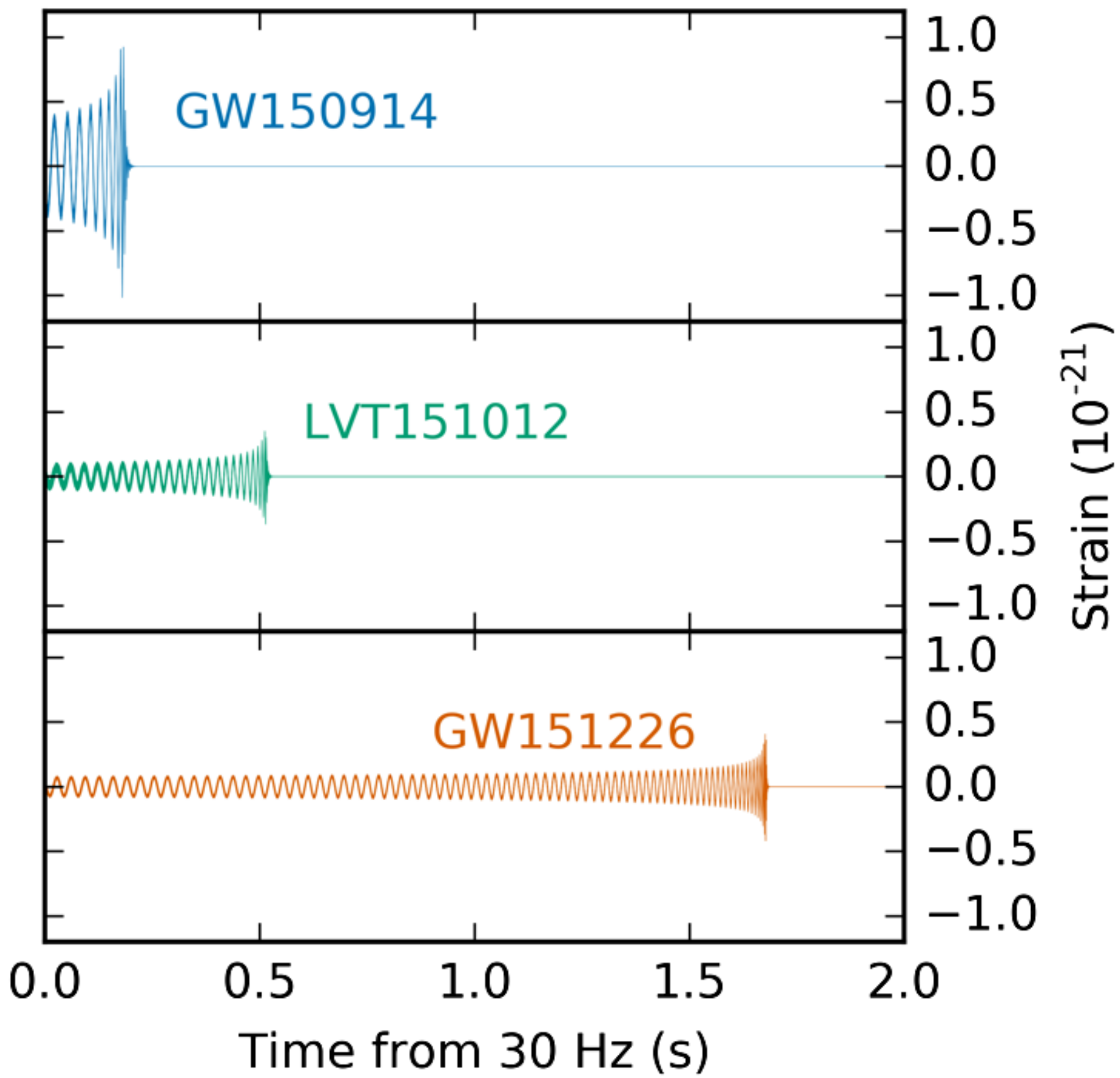


PRL 119 (2017)

Detected Compact Binary Mergers

BBHs in O1

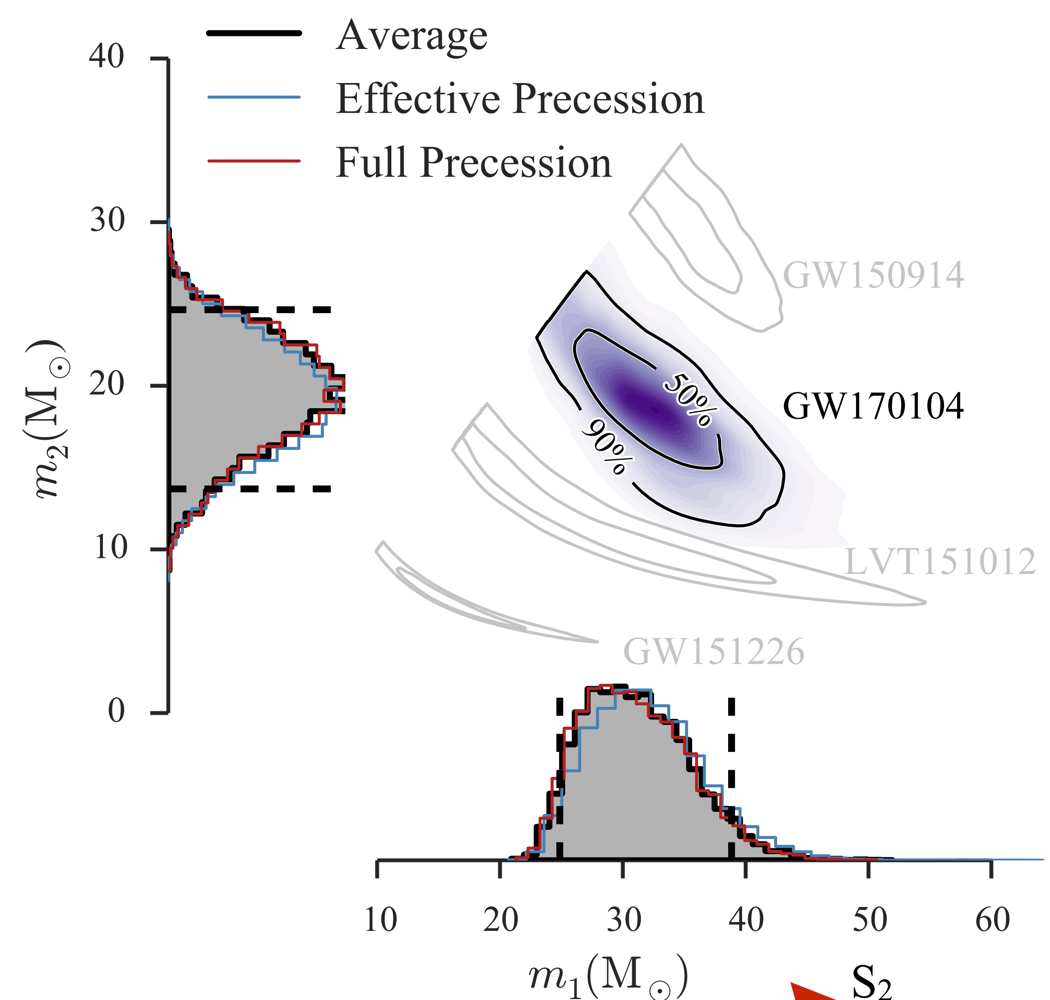
- 2 clear detections and a likely third (LVT)
- Established population of coalescing BBH with component masses up to ~36 $M_\odot$
  - Clear these would form a large fraction of GW detections
- Enabled novel EM followups, population studies, tests of GR, …
- Prompted development of waveforms, NR, …
- But let's focus on O2 events!

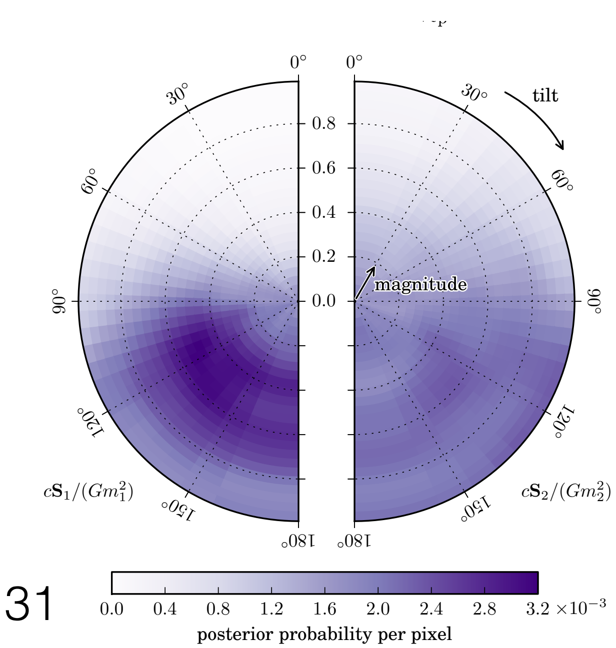


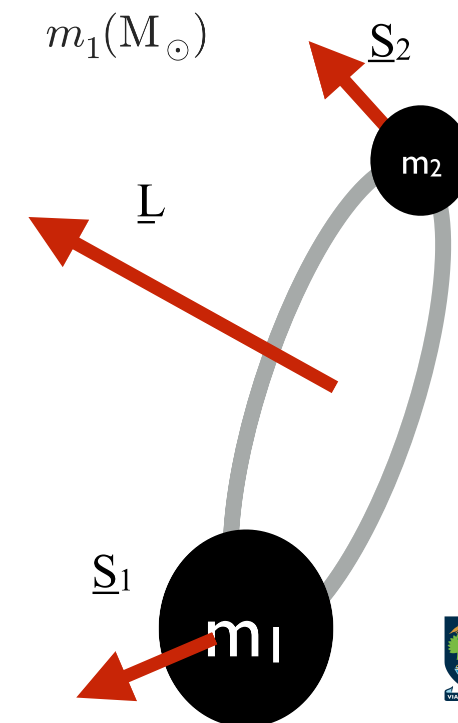
Abbott+ Phys. Rev. X 6, 041015

GW170104

- 50 $M_{\odot}$ total mass
  - sits between masses of O1 events
- 880Mpc (z~0.2)
  - most distant BBH
  - stringent tests of $m_{graviton}$ through dispersion
- Spins likely not positively aligned (although could be non-spinning)
- `LVC 2017 1706.01812`

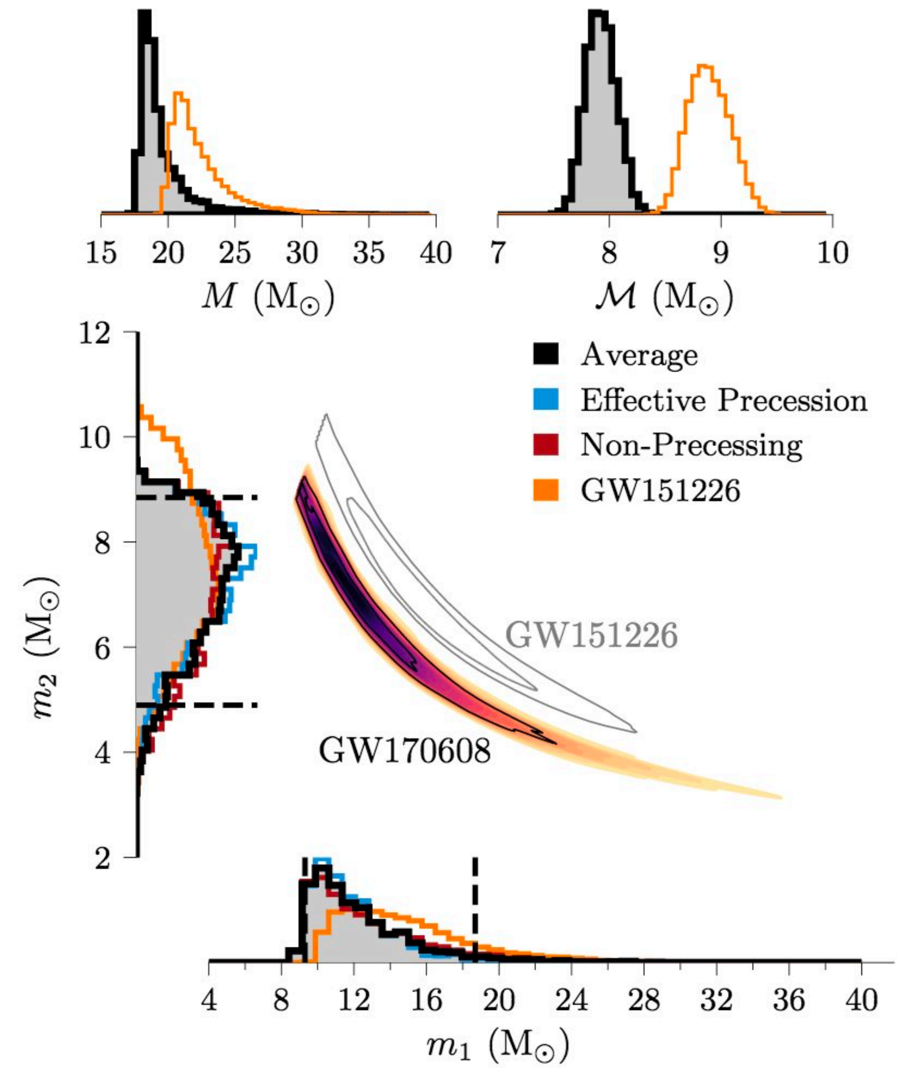


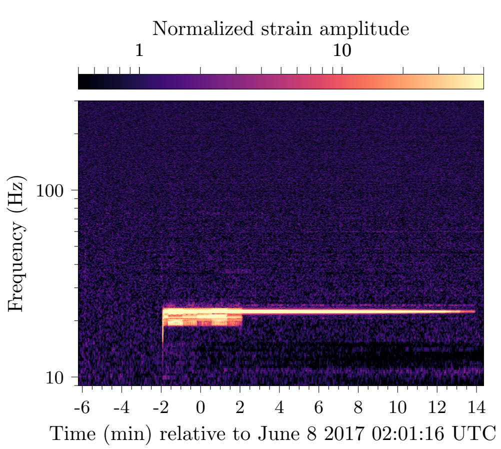


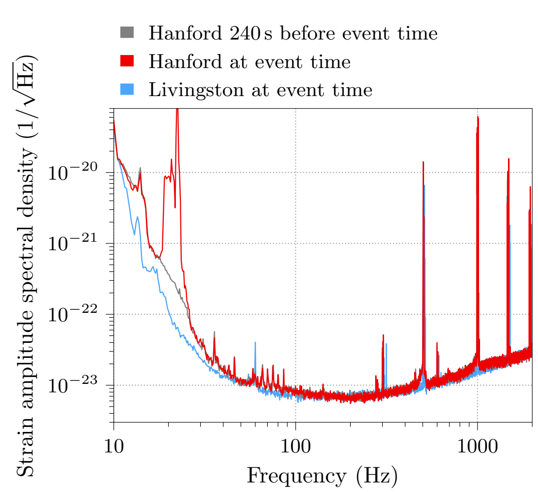


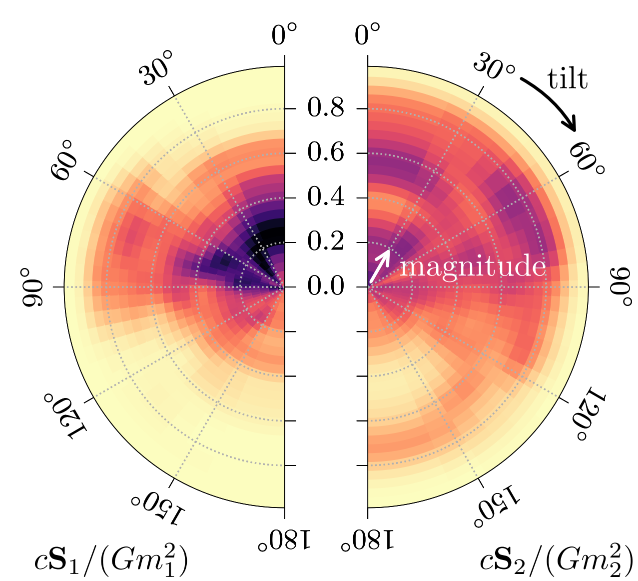
GW170608

- Lightest binary BH yet discovered
  - ~12 + 7 $M_\odot$
  - *Comparable with galactic BH systems known from X-ray observations*
- ~1 M_{sun} radiated as GW energy
- Distance: 340 Mpc (z~0.07)
- Detected during time when Hanford being commissioned
- <u>See arXiv:1711.05578</u> for details

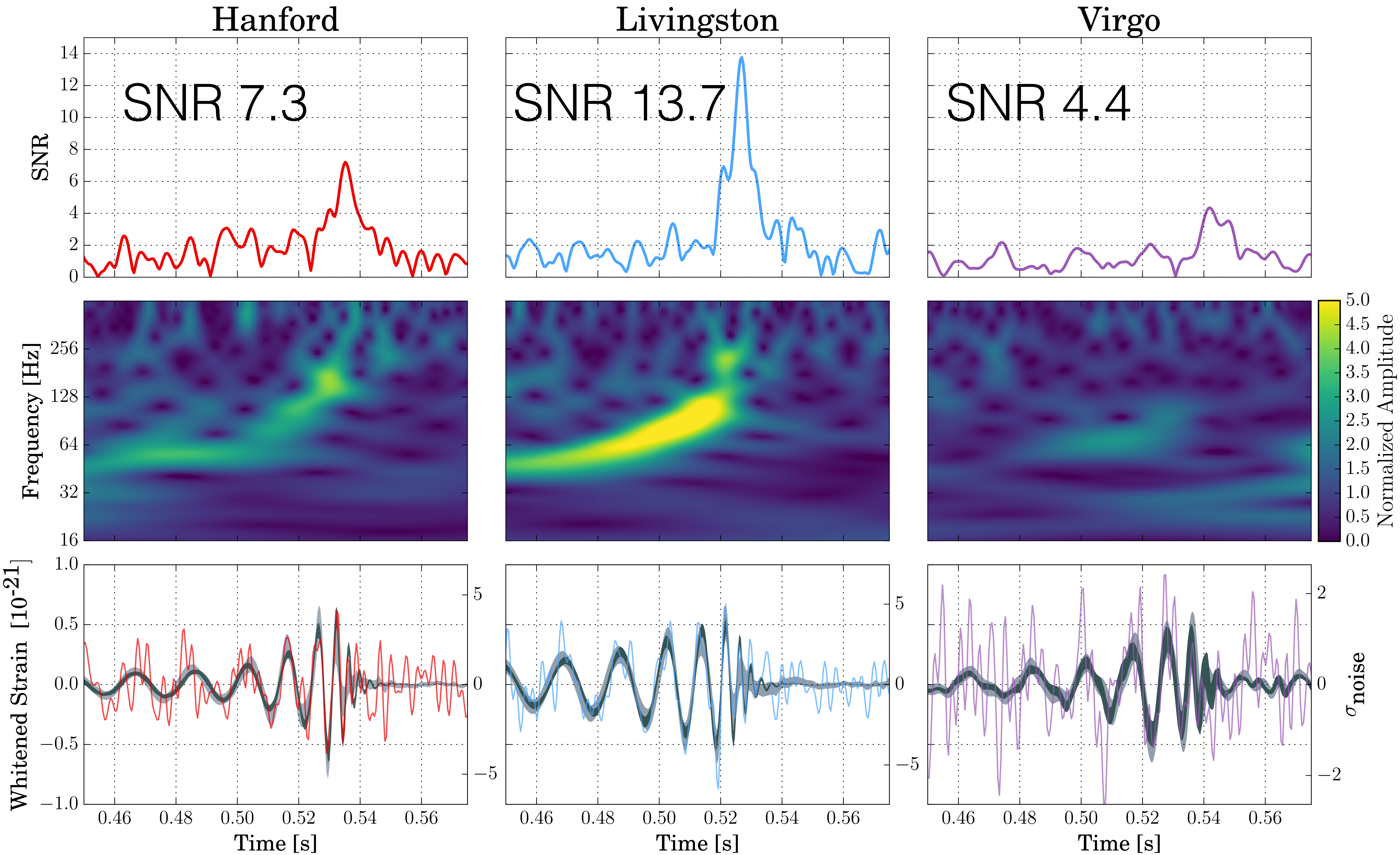






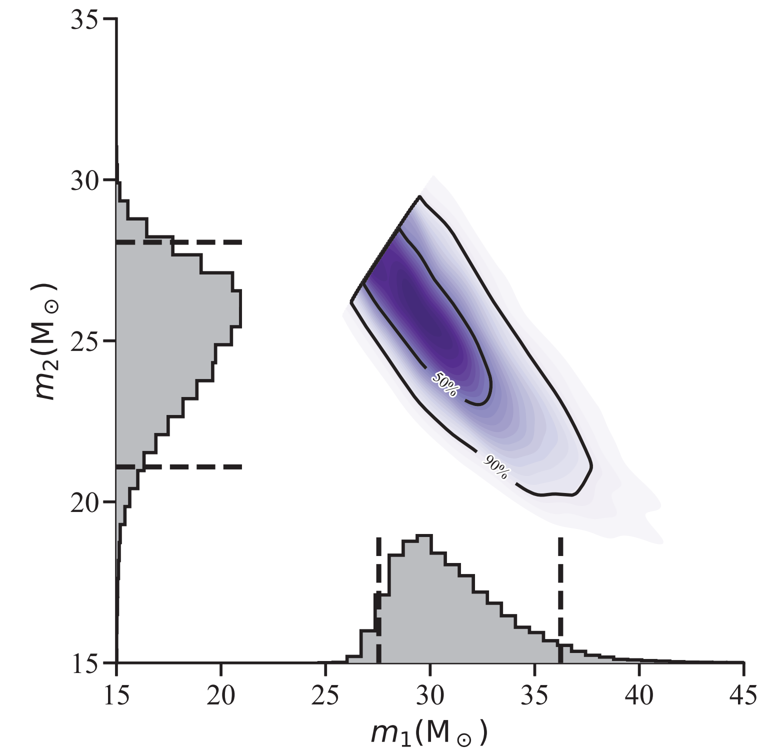


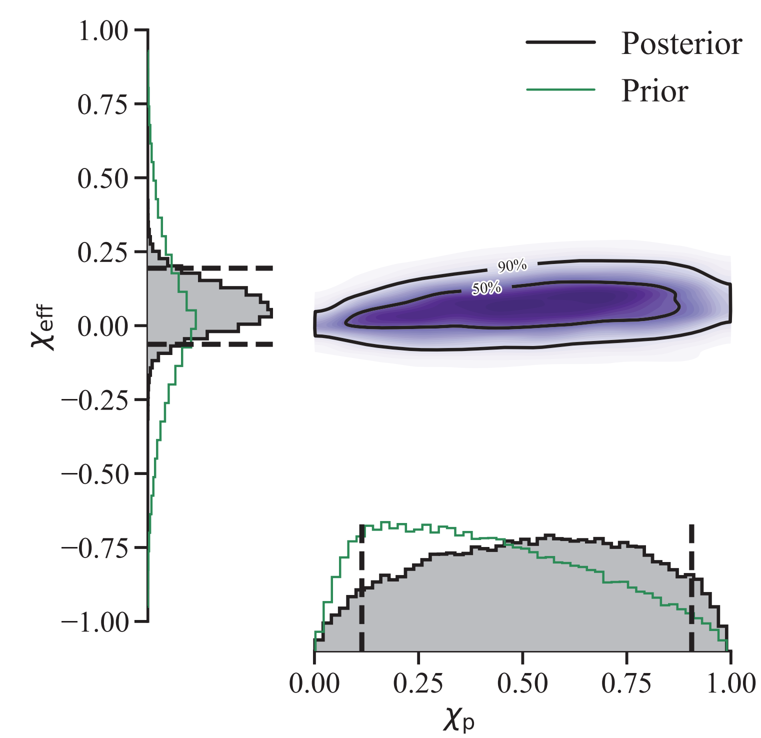
GW170814


Hanford
Livingston
Virgo
SNR 7.3
SNR 13.7
SNR 4.4
SNR
Frequency [Hz]
Normalized Amplitude
Whitened Strain [10^-21]
σ_noise
Time [s]

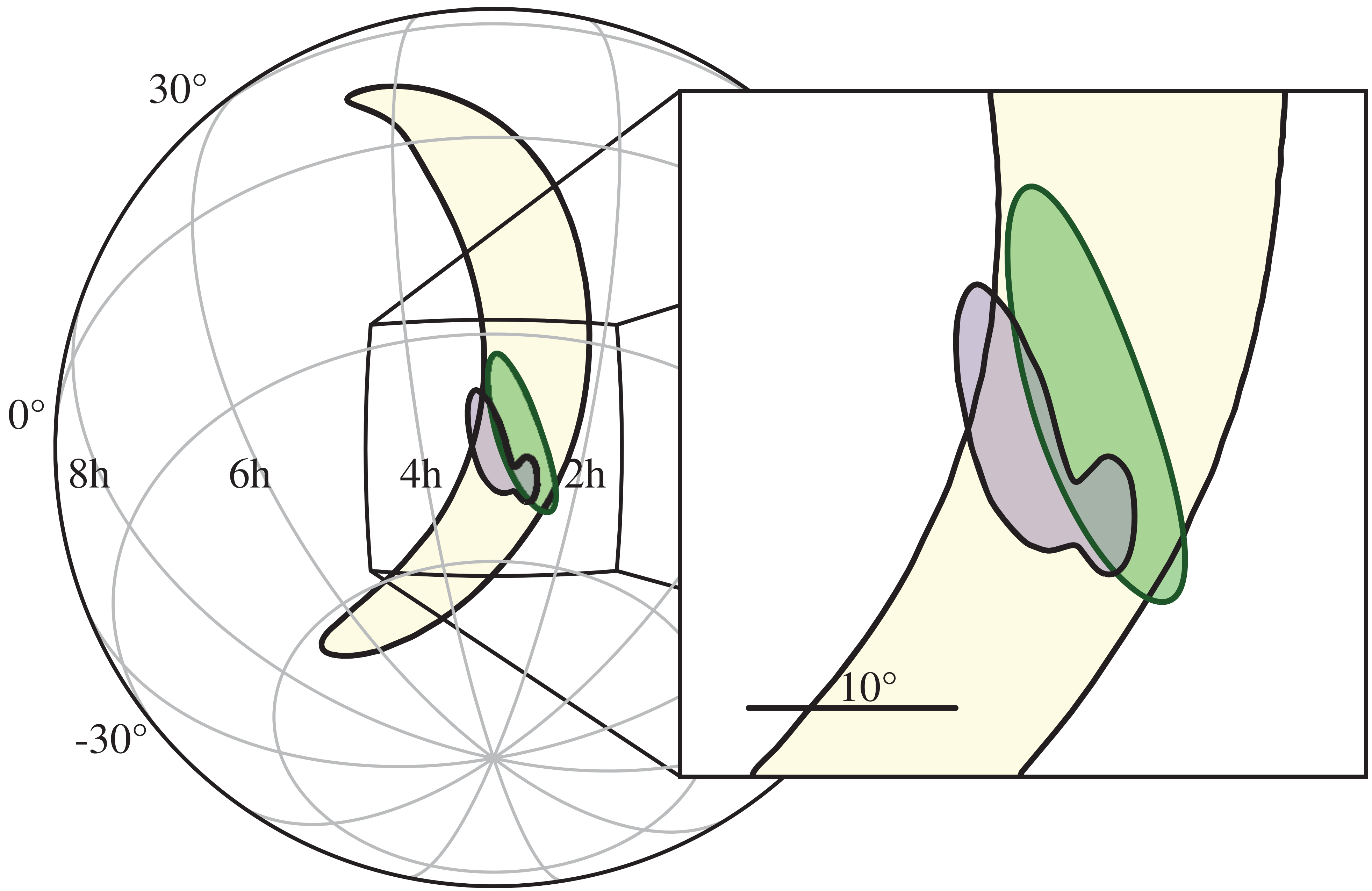
GW170814

- $M_{tot} \sim 56\ M_\odot$
- $d_L \sim$ 540Mpc (z~0.11)
- Poor constraint on spin tilts
- But novel checks of GW polarisation states
- `LVC PRL 119 (2017)`

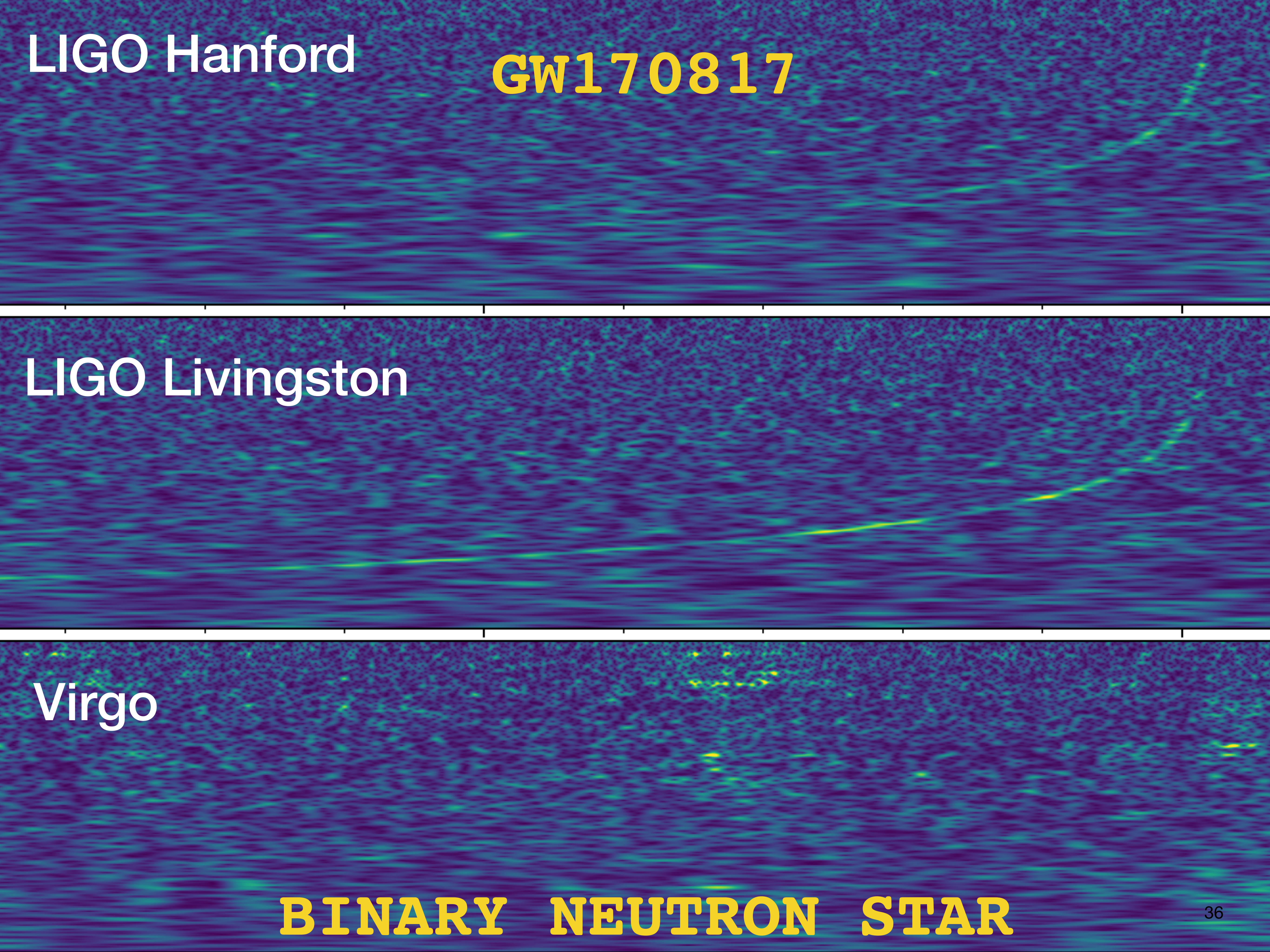




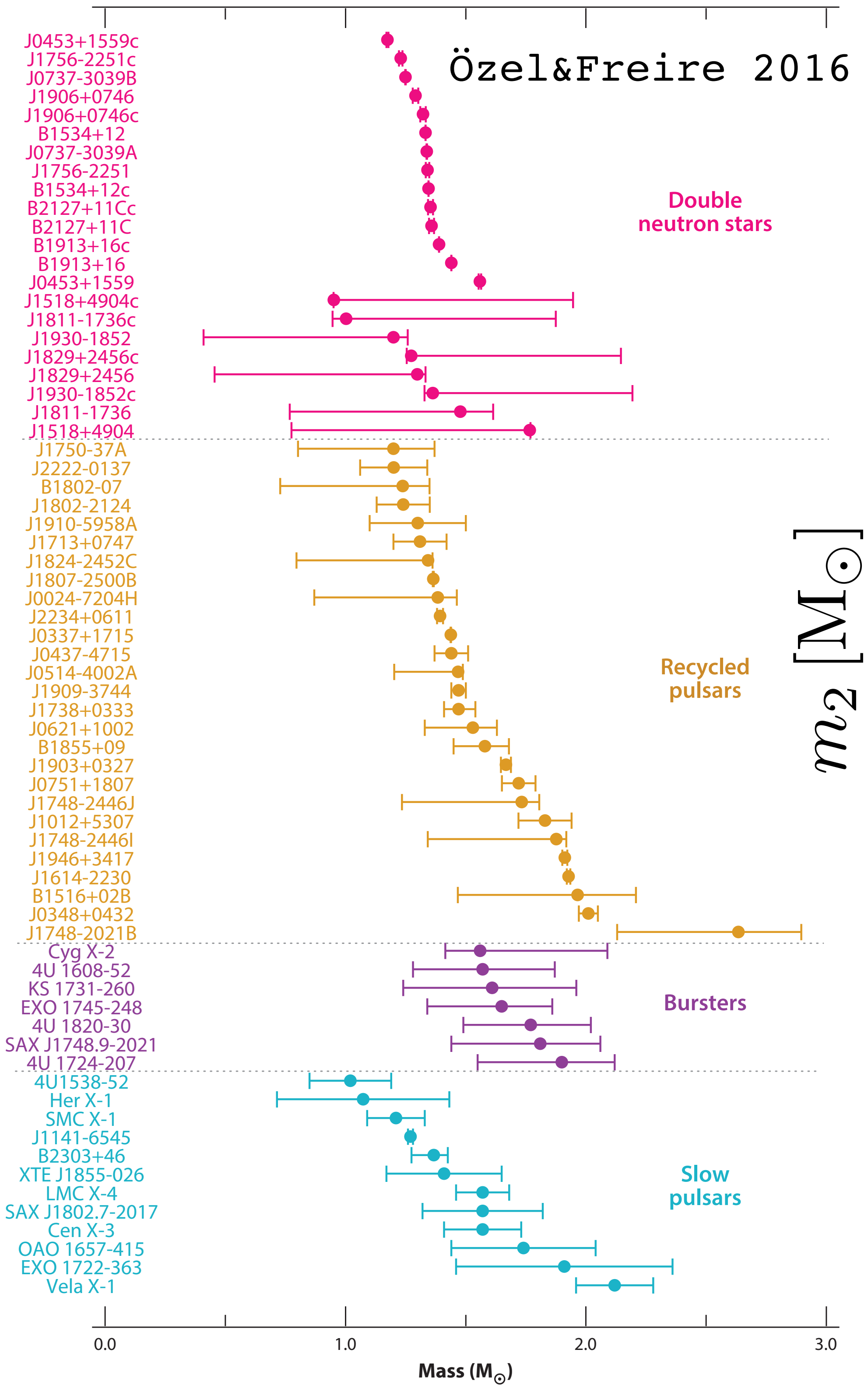
GW170814

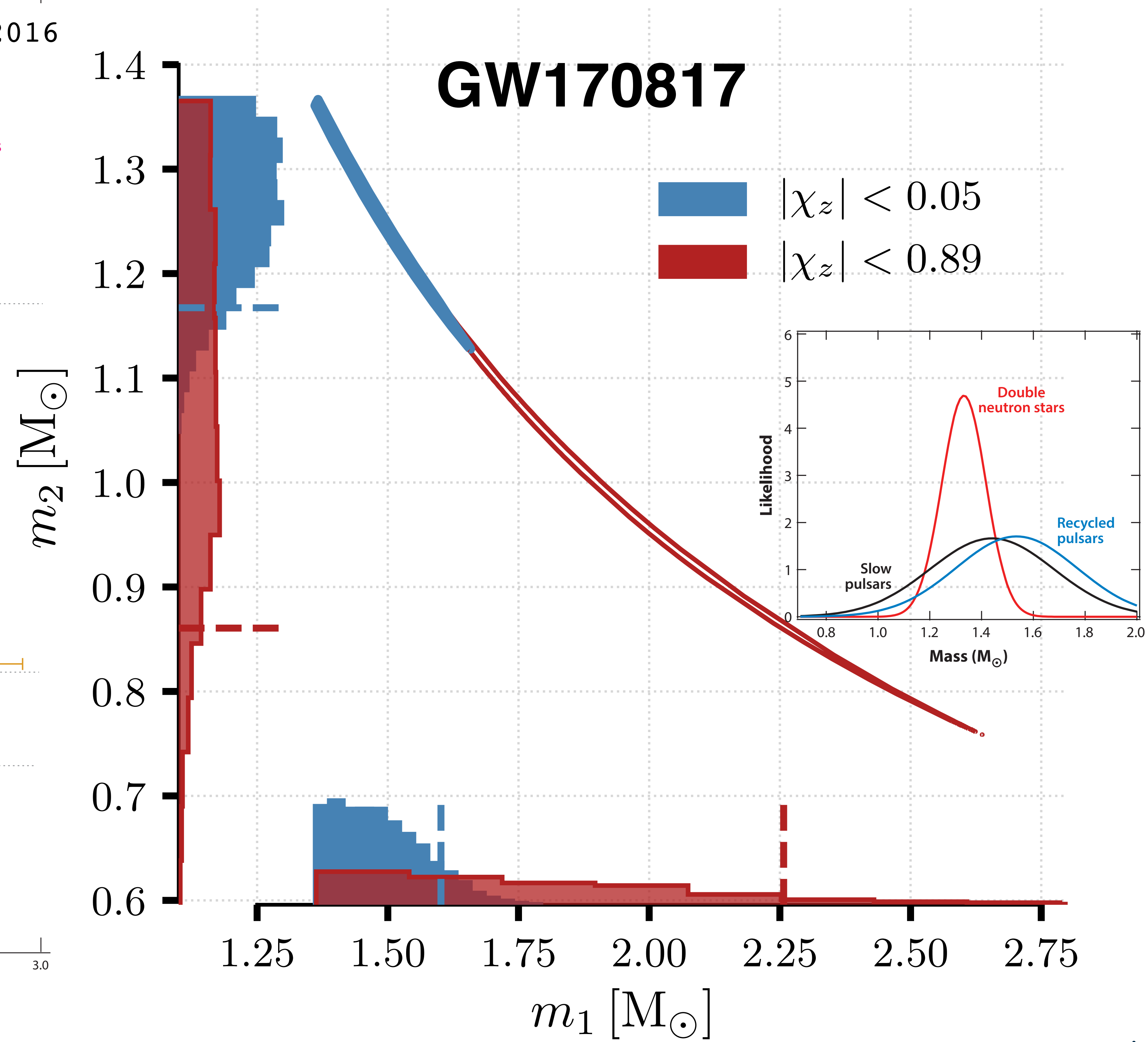


University of Glasgow

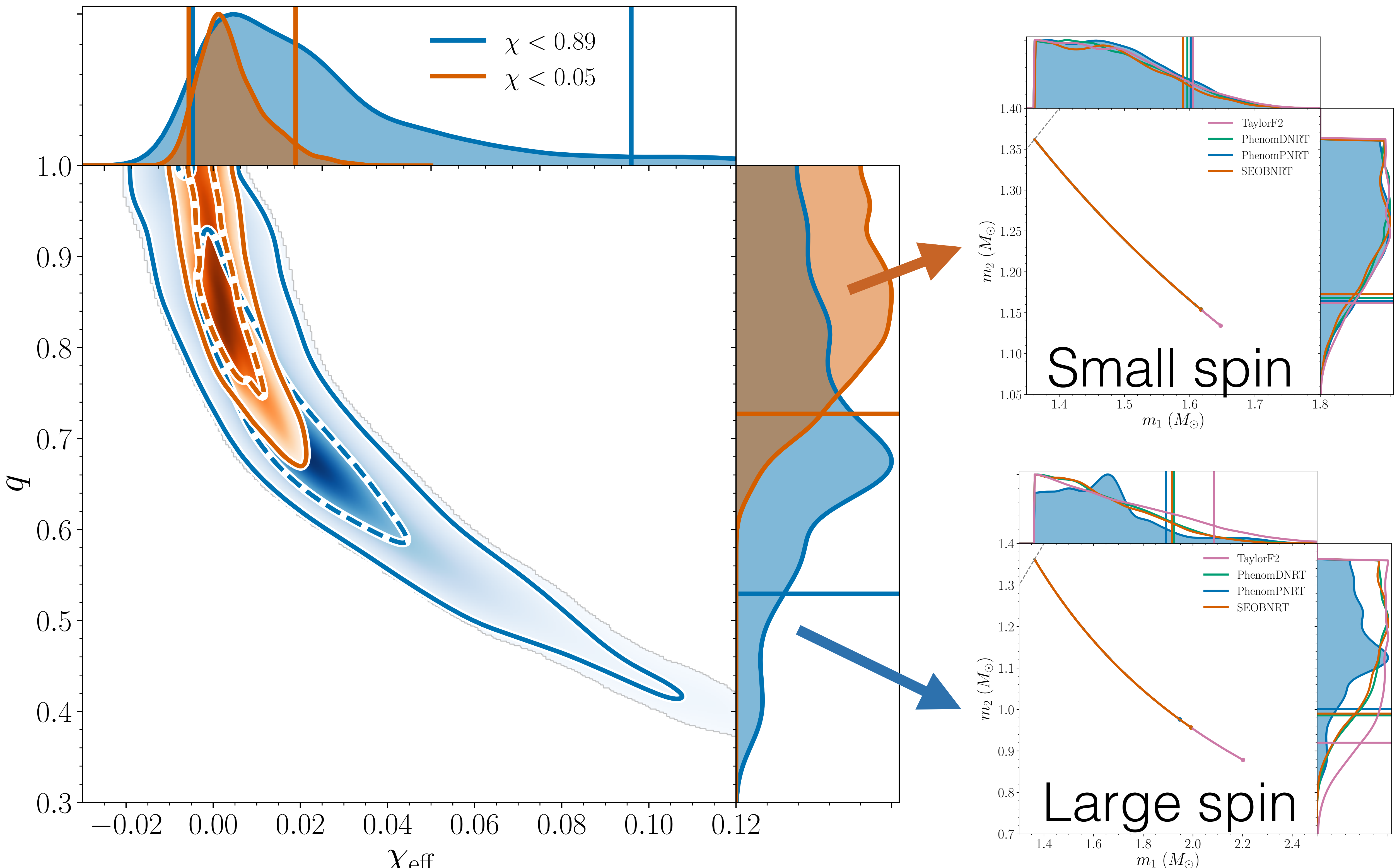

LIGO Hanford
GW170817
LIGO Livingston
Virgo
BINARY NEUTRON STAR

Neutron star masses

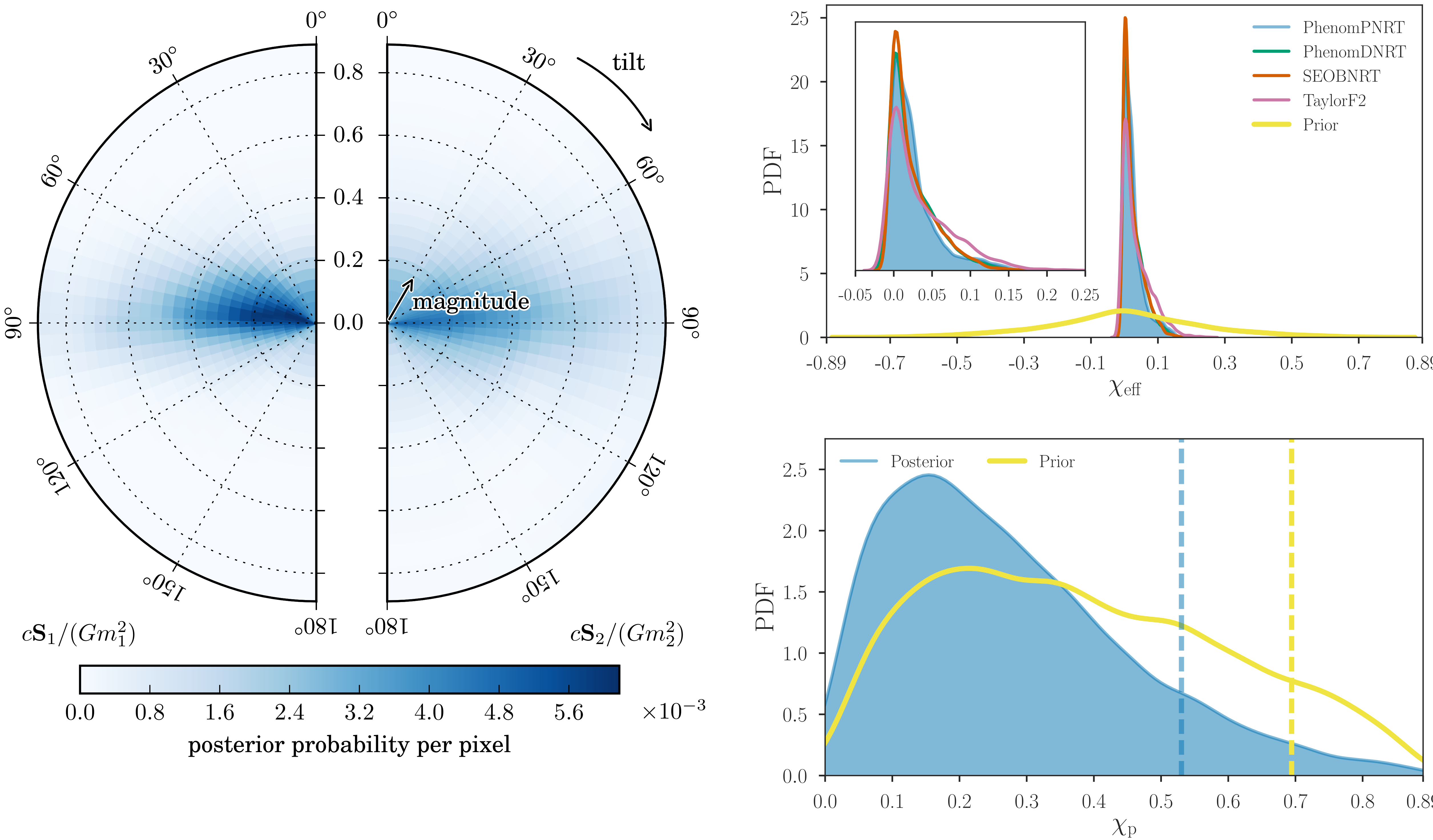




Mass, mass ratio, spin



BNS Spins

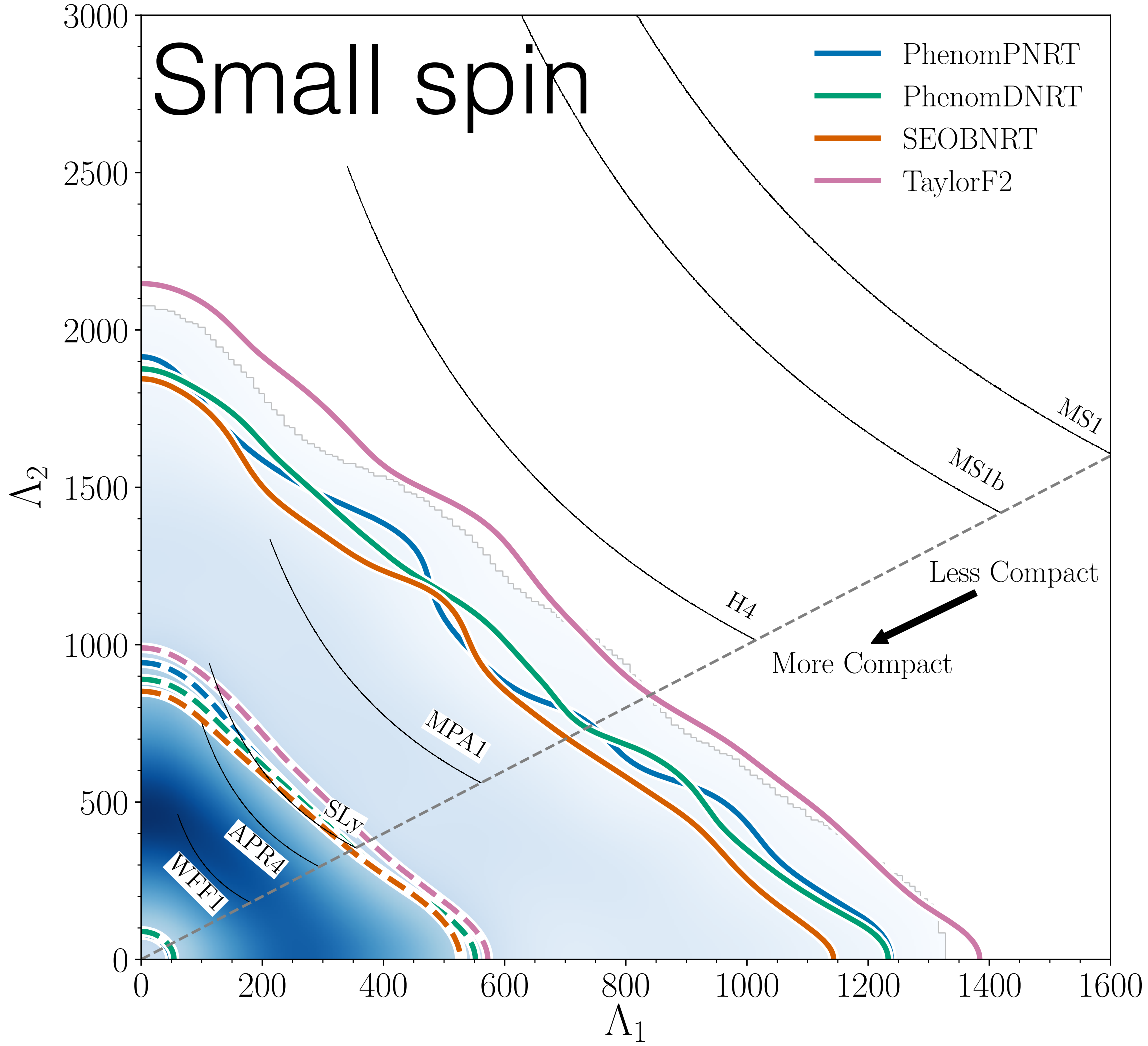


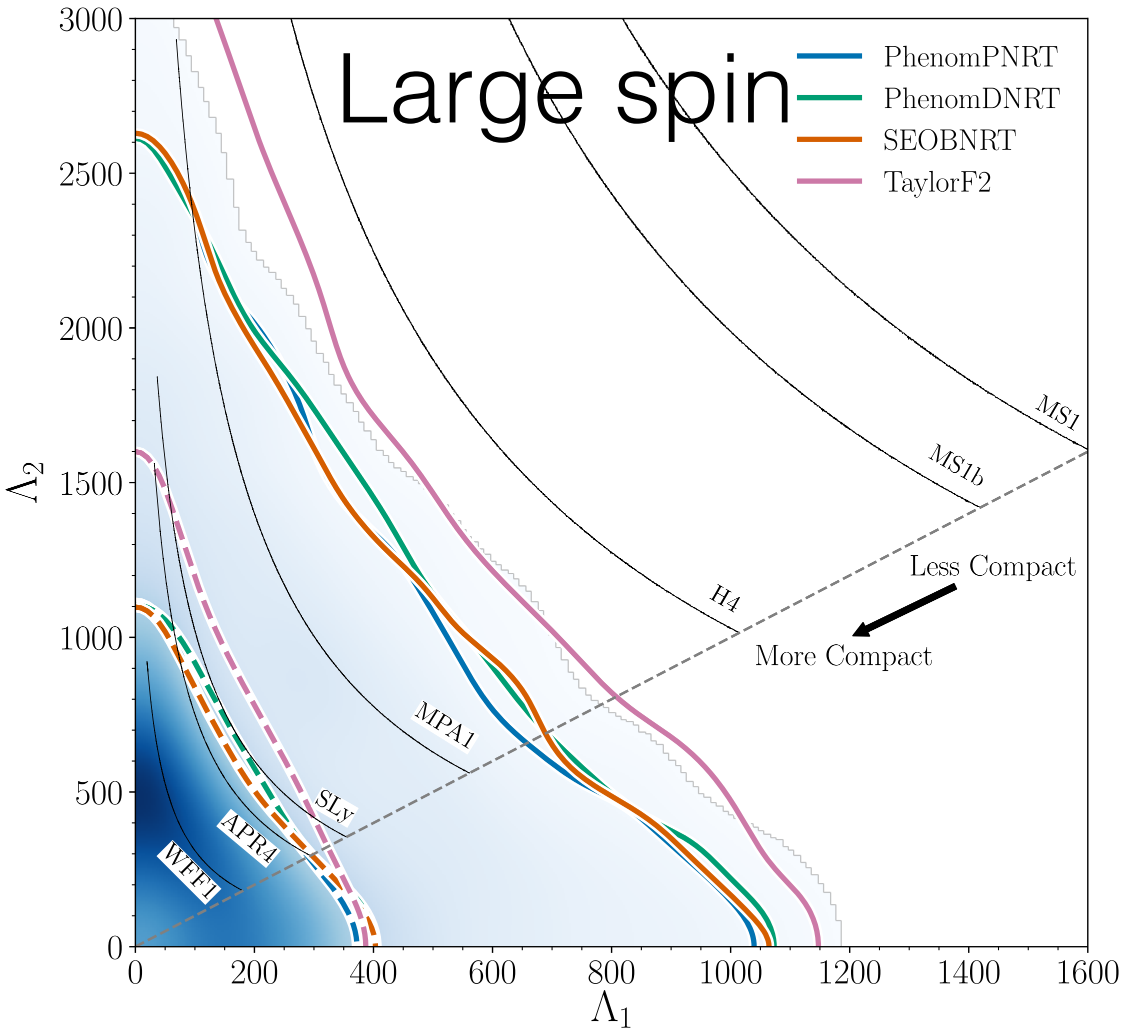
39

arXiv:1805.11579

Tidal deformability

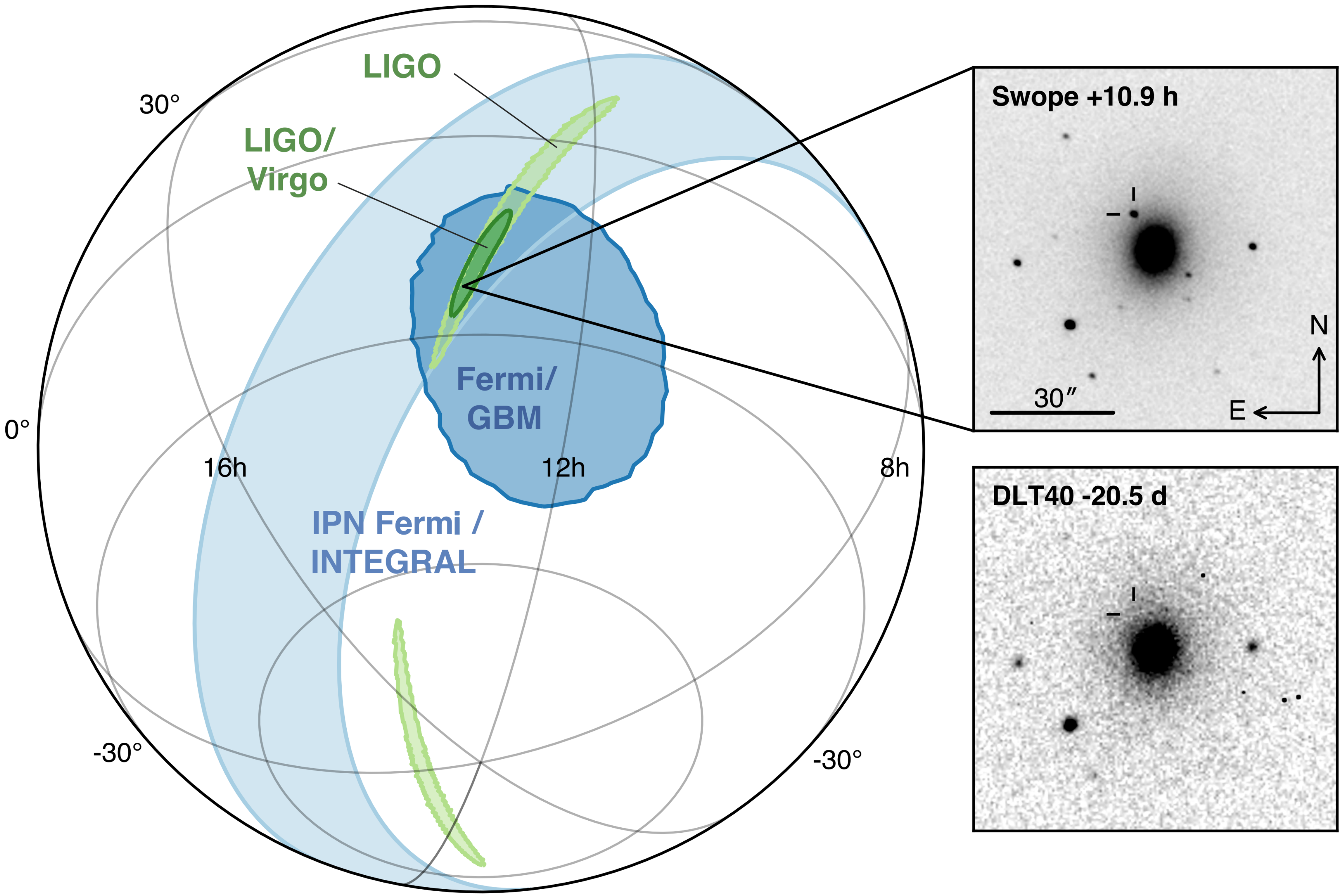
- New analysis with cleaned data, better waveforms [LVC 1805.11579]
- Tighter constraints on tidal deformability
- Spin prior feeds into mass uncertainty -> different masses and compactness results
- See 1805.11581 for further analysis of EOS and compactness





arXiv:1805.11579

Sky localisation



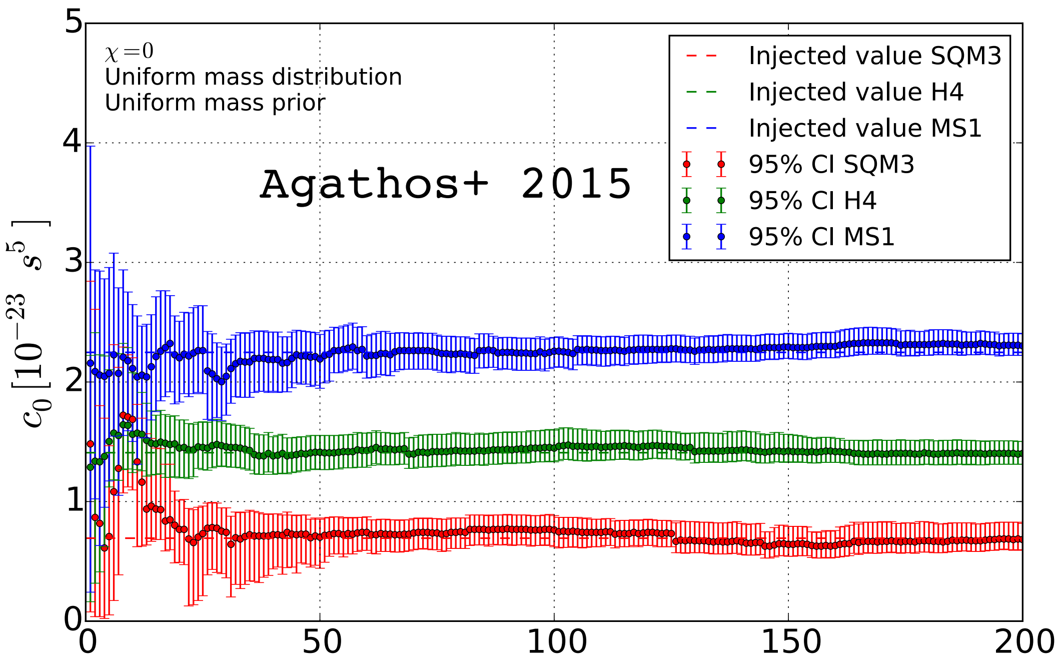
LSC VIRGO

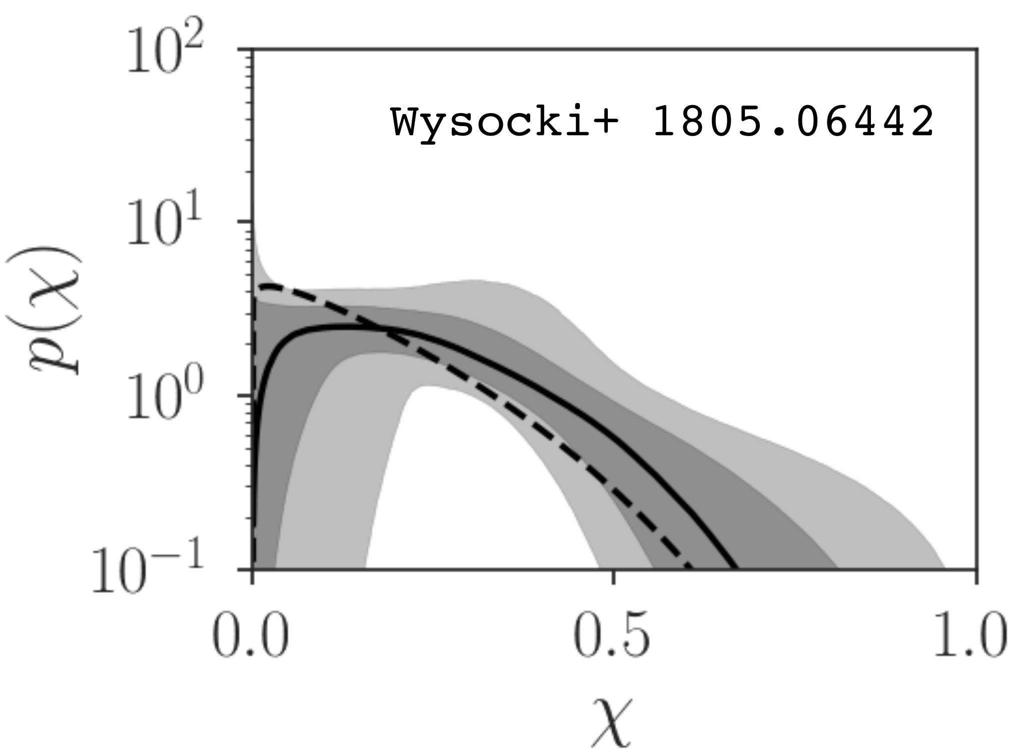
University of Glasgow

Hierarchical models: combining events

Hierarchical Models

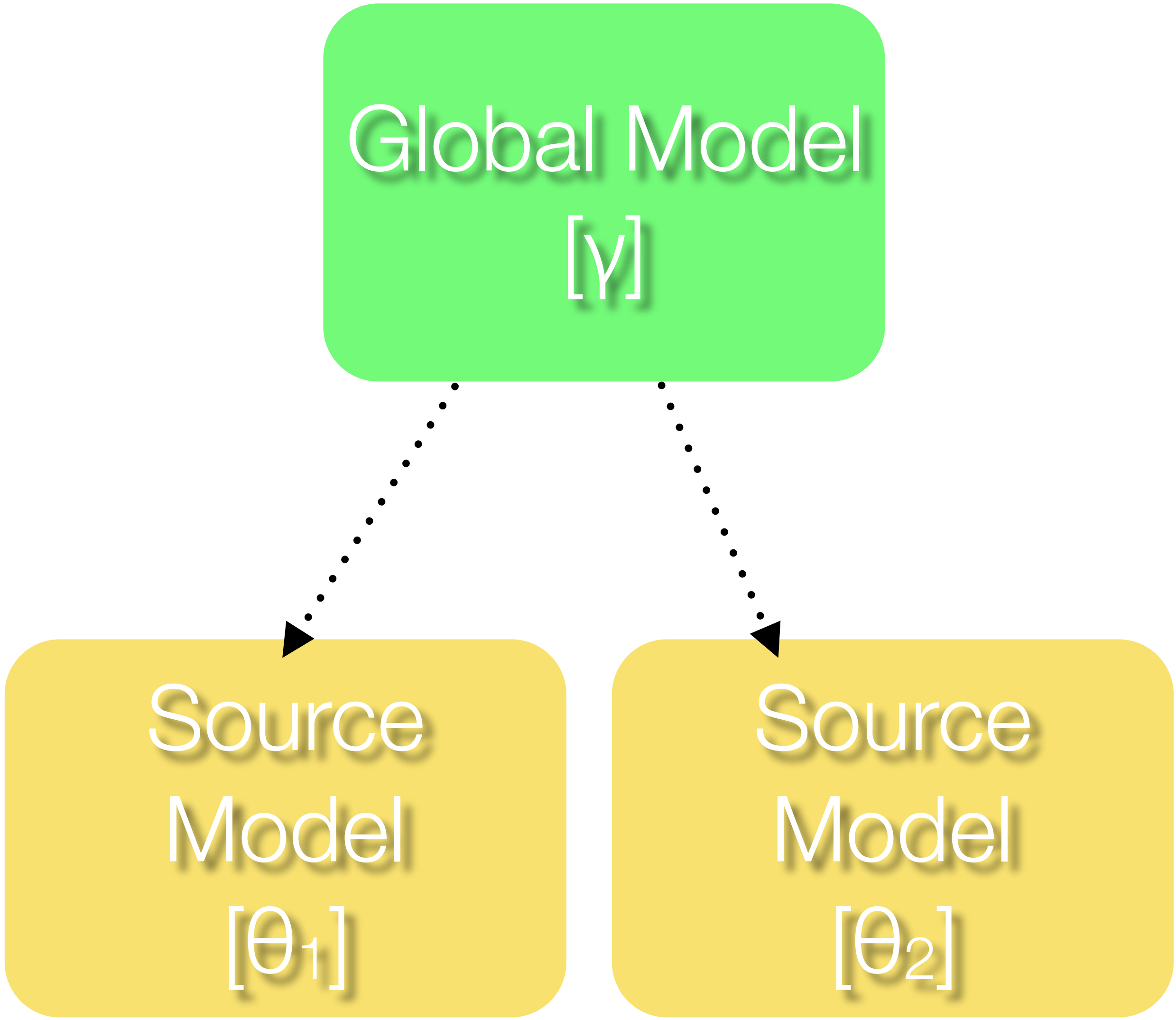
- Inference from multiple events about parameters affecting all events, e.g.
  - Common parameters: Tests of General Relativity, Neutron Star EOS
  - Population parameters: Rate, BH mass function, spin distributions, …
- Selection effects are important!





Hierarchical Models

- Goal: Extract information about some **population parameter** γ from *N* detected events with data *d* and nuisance parameters θ



- $$p(\gamma|N, \{d_1, d_2, ...\}) \propto p(\gamma|I)p(N|\gamma)\prod_i^N p(d_i|\gamma, \text{detection})$$

Selection effects

- In general, the number of detected sources *N* can be affected by the source properties
  - e.g. Large masses -> larger distances
- We are interested in the astrophysical distribution not the observed distribution
  - Must account for selection effects!

$$p(N|\gamma) = p(N|\hat{N})p(\hat{N}|\gamma) = \frac{\hat{N}(\gamma)^N e^{-\hat{N}(\gamma)}}{N!}$$

Poisson statistics on observed number *N*

$$\hat{N} = \int \frac{\partial^k \hat{N}(\gamma)}{\partial \vec{\theta}} d\vec{\theta}$$

$$= \int \frac{\partial^k N_{astro}(\gamma)}{\partial \vec{\theta}} p(\text{detection}|\theta) d\theta$$

Observed distribution is astrophysical distribution filtered by detection probability

LSC VIRGO

University of Glasgow

Use of single event samples

- Global event likelihood factorises into single-event likelihoods

$$p(\{d_1, d_2, ...\}|\gamma) = \prod_i^N p(d_i|\gamma)$$

- Each event has nuisance parameters θ which may depend on the population parameters γ

$$p(d_i|\gamma) = \int p(d_i|\theta)p(\theta|\gamma)d\theta$$

- Can use samples from *likelihood* to evaluate the integral

$$p(d_i|\gamma) = \langle p(\theta|\gamma)\rangle_{p(d_i|\theta)}$$

$$\approx \frac{1}{N_{\rm samps}} \sum_i^{N_{\rm samps}} p(\theta_i|\gamma)$$

LSC VIRGO University of Glasgow

Example: Anisotropy of BBHs?

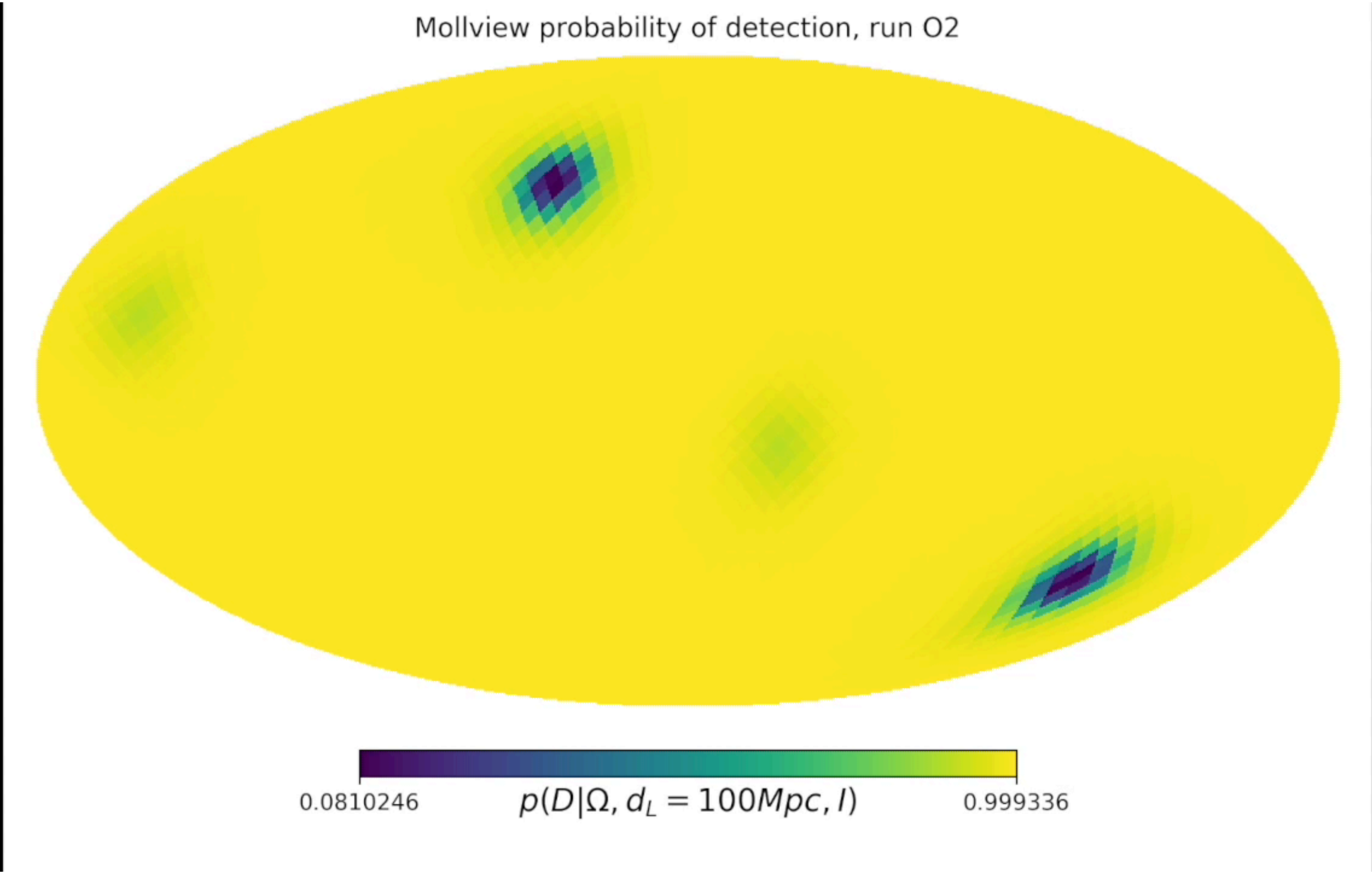
- Question: How are BBHs distributed over the sky?
  - Isotropic: $p(\Omega|I)=1/(4\pi)$
  - Anisotropic: $p(\Omega \in \text{pixel}_i \,|\, a_i) \propto a_i$
- Need to know:
  - Source locations: PE samples
  - Selection function: Detector sensitivity (time-dependent)

$$P\left(x_j \,|\, \vec{a},\, \mathrm{D},\, I\right) = P\left(\mathrm{D} \,|\, x_j,\, \vec{a},\, I\right) \frac{\int d\Omega \int p\left(x_j \,|\, \Omega, \mathcal{D},\, I\right) p\left(\Omega,\, \mathcal{D} \,|\, \vec{a},\, I\right) d\mathcal{D}}{\int d\Omega \int p\left(\mathrm{D} \,|\, \Omega, \mathcal{D},\, I\right) p\left(\Omega,\, \mathcal{D} \,|\, \vec{a},\, I\right) d\mathcal{D}}$$

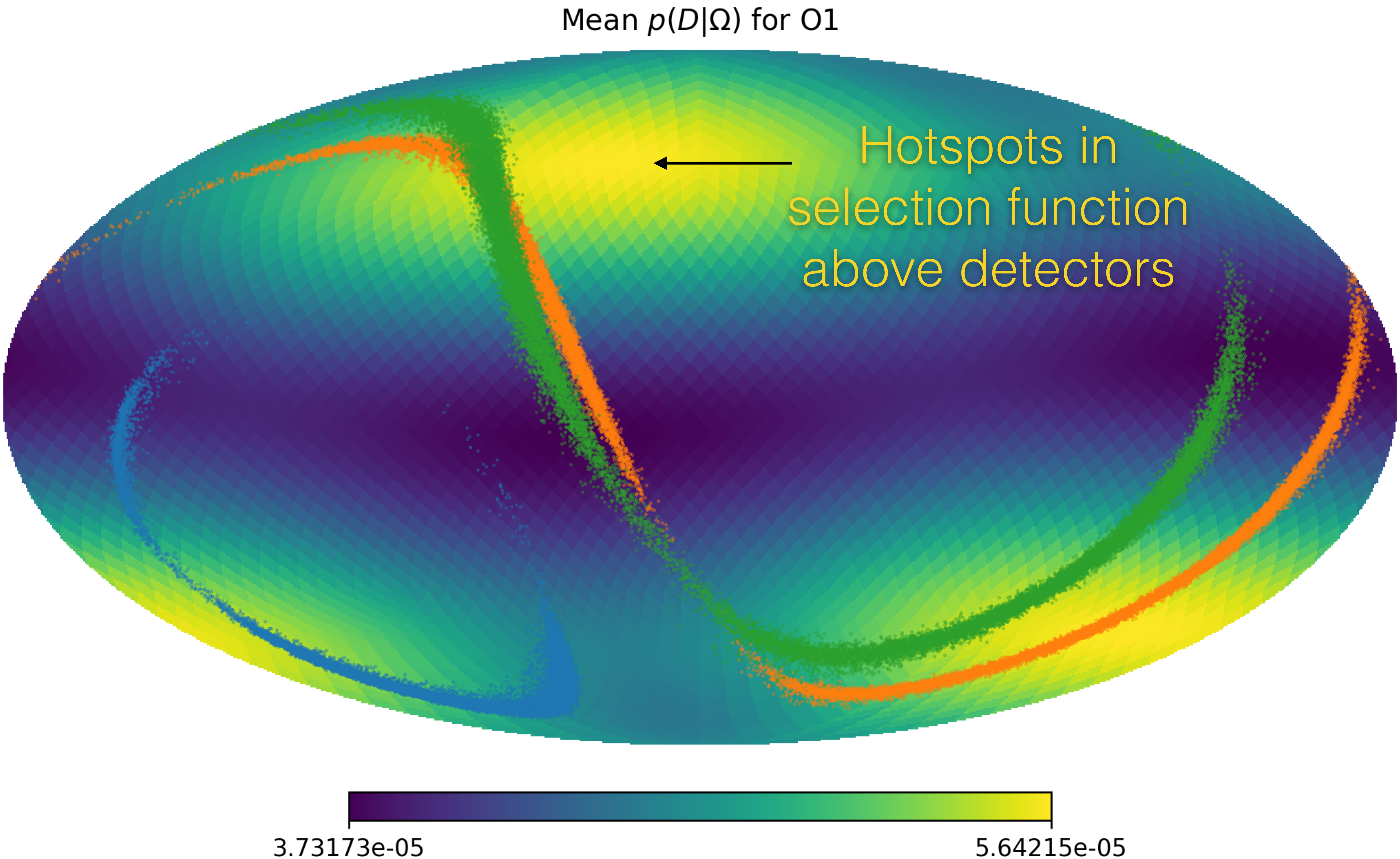
$$\approx \frac{\frac{1}{N}\sum_j^N p\left(\Omega_j,\, \mathcal{D}_j \,|\, \vec{a},\, I\right)}{\int p\left(\mathrm{D} \,|\, \Omega,\, I\right) p\left(\Omega \,|\, \vec{a},\, I\right) d\Omega},$$

$$\begin{aligned}\widehat{N}\left(\vec{a}\right) &= N_{\text{merg}} \iint \frac{\partial^2 N_{\text{obs}}}{\partial\Omega\partial\mathcal{D}} d\mathcal{D} d\Omega \\ &= RVT \iint p\left(\mathrm{D} \,|\, \Omega,\, \mathcal{D},\, I\right) p\left(\Omega,\, \mathcal{D} \,|\, \vec{a},\, I\right) d\mathcal{D} d\Omega \\ &= RVT \int p\left(\mathrm{D} \,|\, \Omega,\, I\right) p\left(\Omega \,|\, \vec{a},\, I\right) d\Omega,\end{aligned}$$

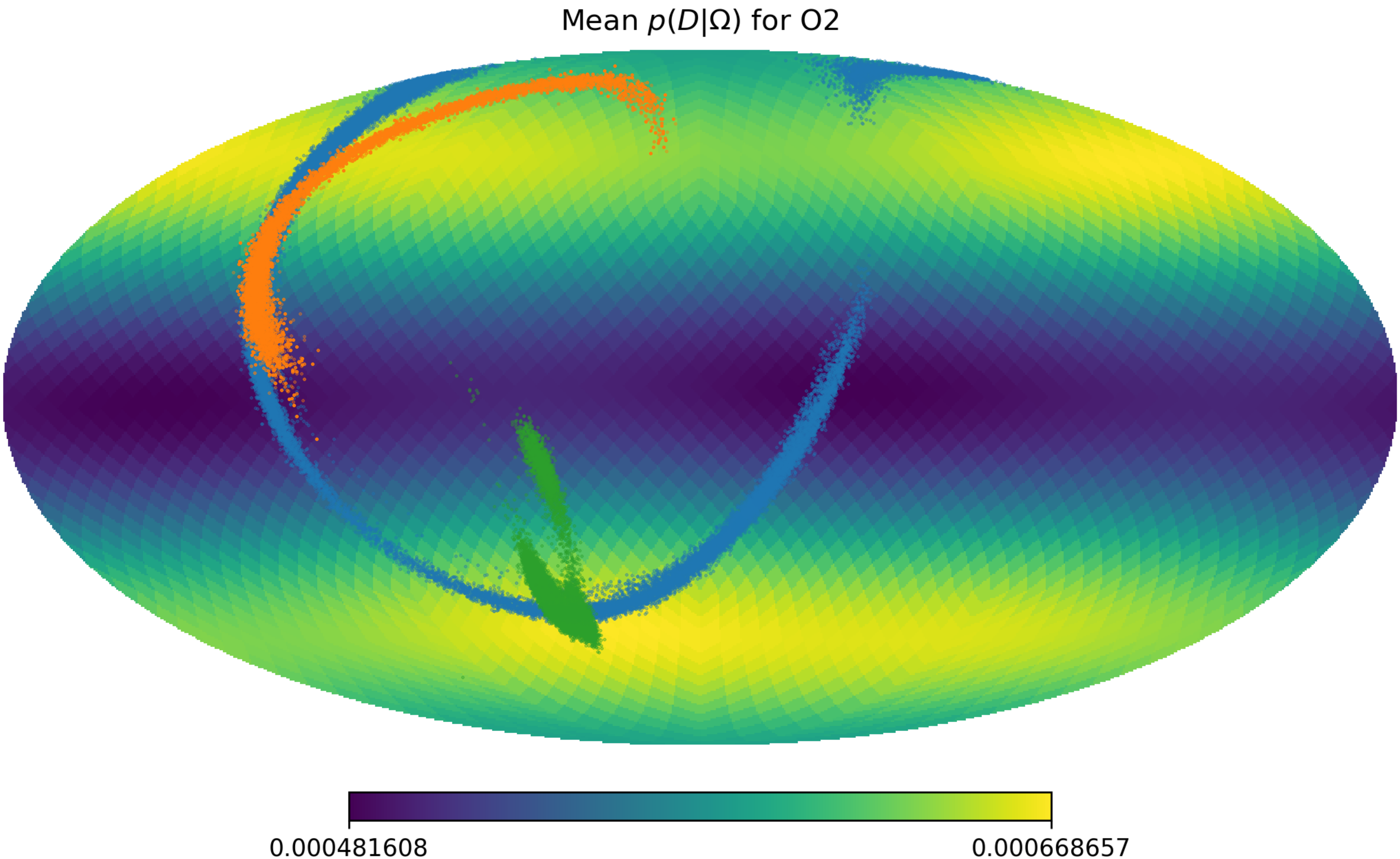
O2 selection function



O1 data

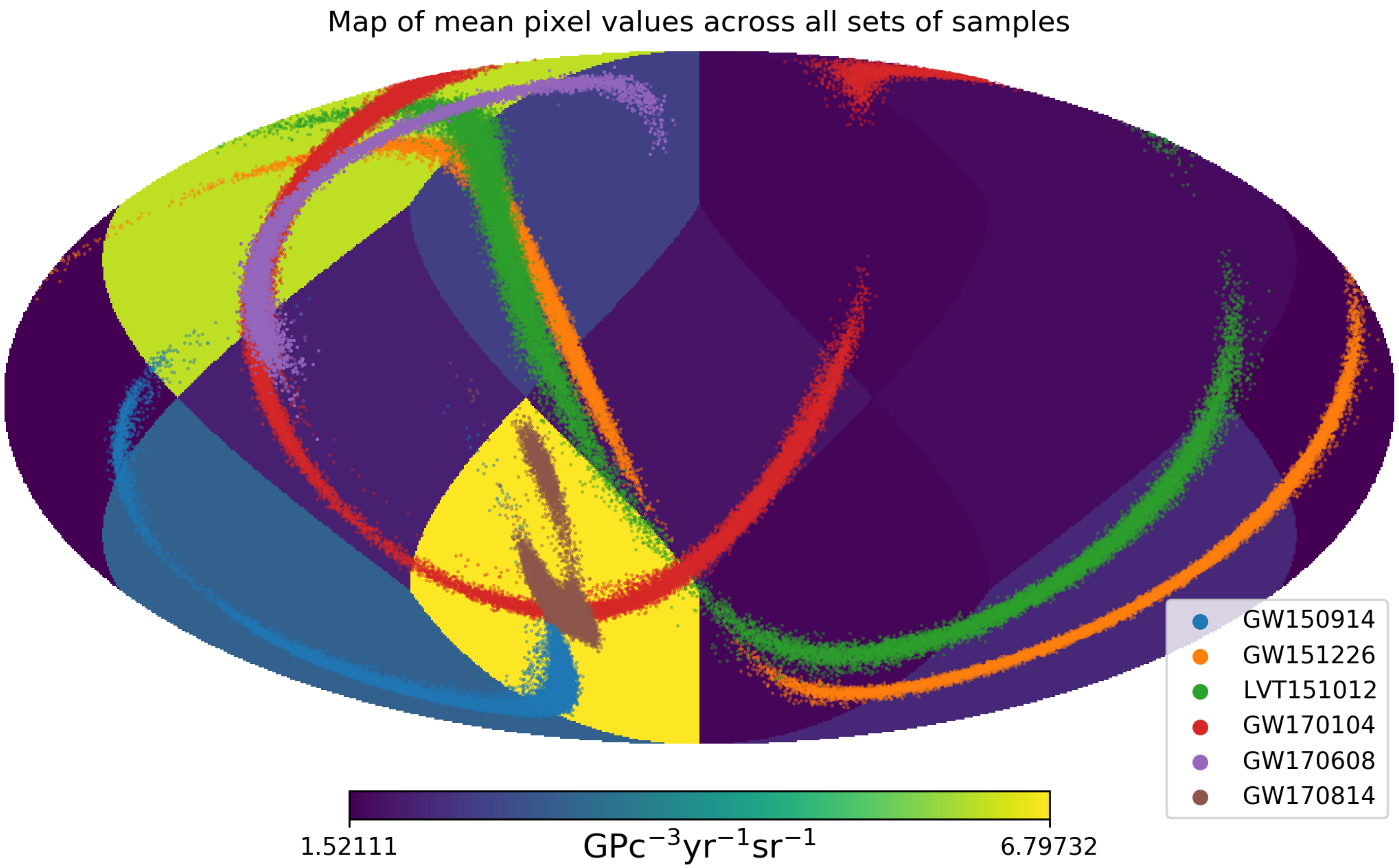

Mean $p(D|\Omega)$ for O1
Hotspots in selection function above detectors
3.73173e-05
5.64215e-05

O2 data



Results

Posterior on pixel intensity



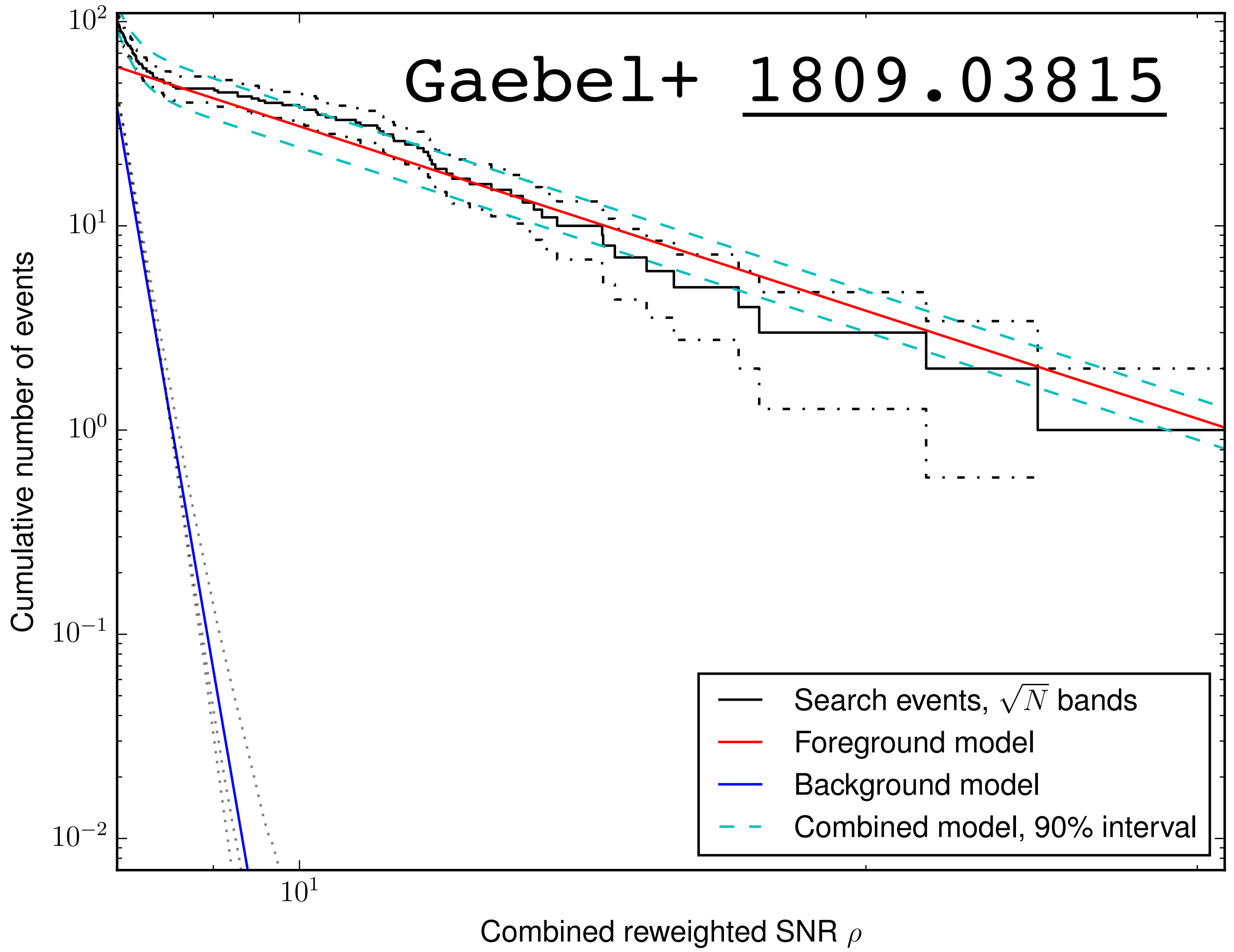
	log Z
Isotropic	-44.6
Anisotropic	-56.1
log Bayes Factor ISO/ANISO	11.5

Isotropic:Pixel = 100,000:1

LSC VIRGO University of Glasgow

Mass distribution with background

- Can we distinguish two populations with a hierarchical model?
  - Astrophysical distribution (F) + noise triggers (B)
  - Use SNR ρ and mass posterior samples m



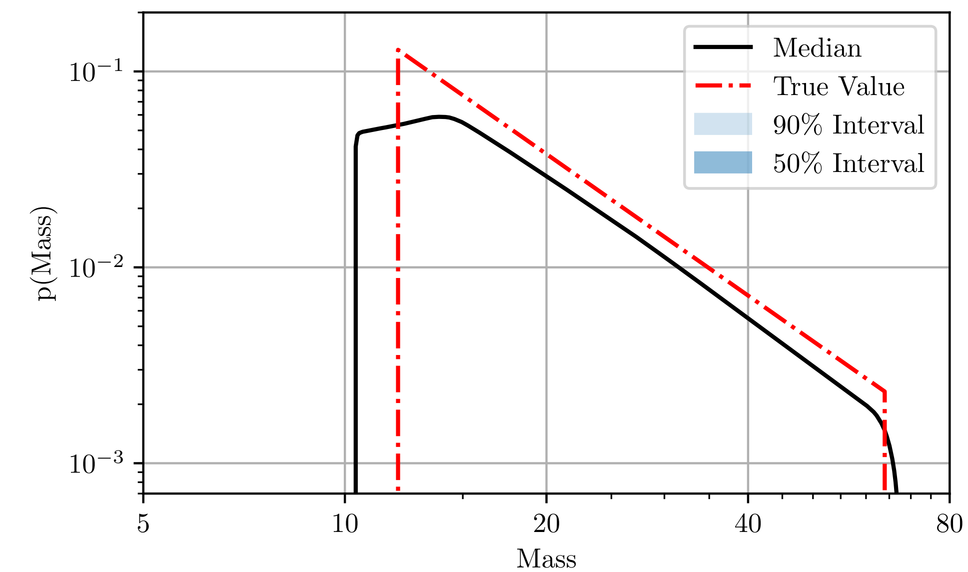
- Mixture model sums components in likelihood for each event

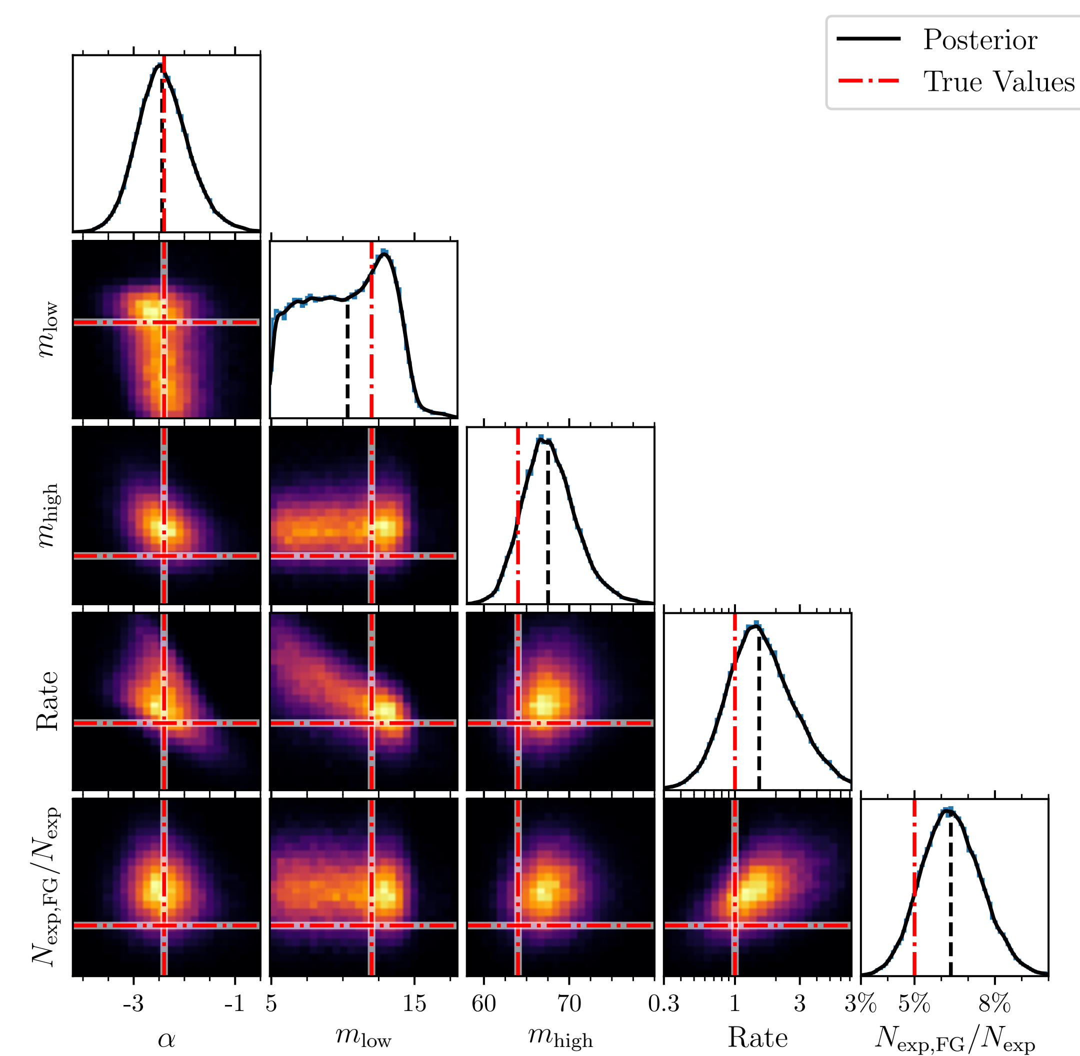
$$p(\rho_i, \vec{m}_i | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}, \theta_{\mathrm{F}}, \theta_{\mathrm{B}}) = p(\rho_i, \vec{m}_i | \theta_{\mathrm{F}}, \eta_i = \mathrm{F}) p(\eta_i = \mathrm{F} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}}) + p(\rho_i, \vec{m}_i | \theta_{\mathrm{B}}, \eta_i = \mathrm{B}) p(\eta_i = \mathrm{B} | N_{\mathrm{F,exp}}, N_{\mathrm{B,exp}})$$

LSC VIRGO University of Glasgow

Truncated power law mass distribution

- Astrophysical mass distribution modelled as power law with min, max cutoff and unknown slope
- Selection function ∝M³
- Estimate rate, slope, cutoffs in presence of noise background



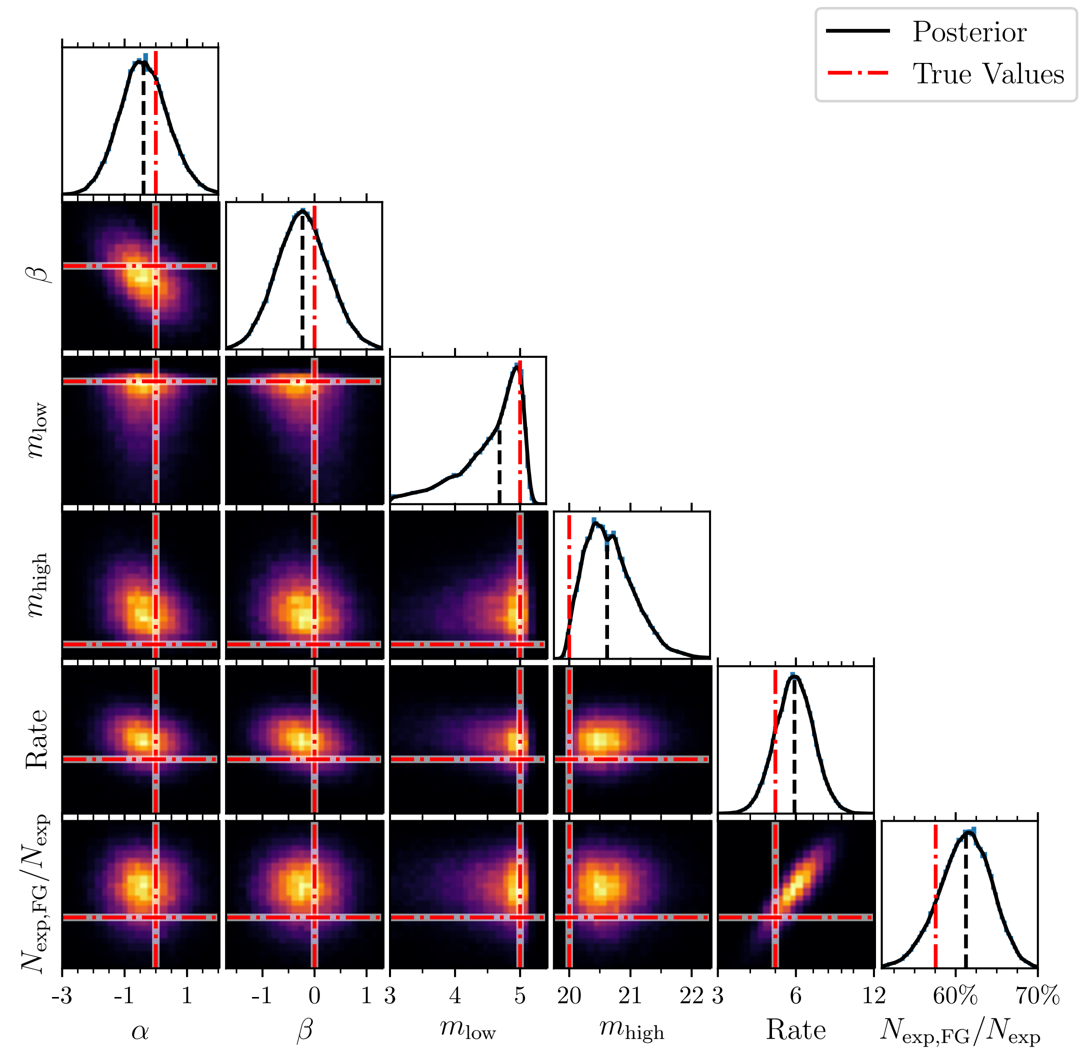


Application to realistic data (ER4)

- As a more realistic case, consider ER4 data with known set of BBH injections
- Include two mass parameters, use real PE posterior samples
- Truncated dual power law foreground mass model

$$p(m_1, m_2|\theta_F, \eta = F) \propto \begin{cases} m_1^\alpha m_2^\beta & \text{if } m_{low} < m_2 \leq m_1 < m_{high} \\ 0 & \text{else} \end{cases}$$

- Background power law in SNR and rate estimated from time-slides, mass from 2D fit to timeslide trigger masses
- Estimation performs well despite imperfect assumptions - could be applied to O1 and O2

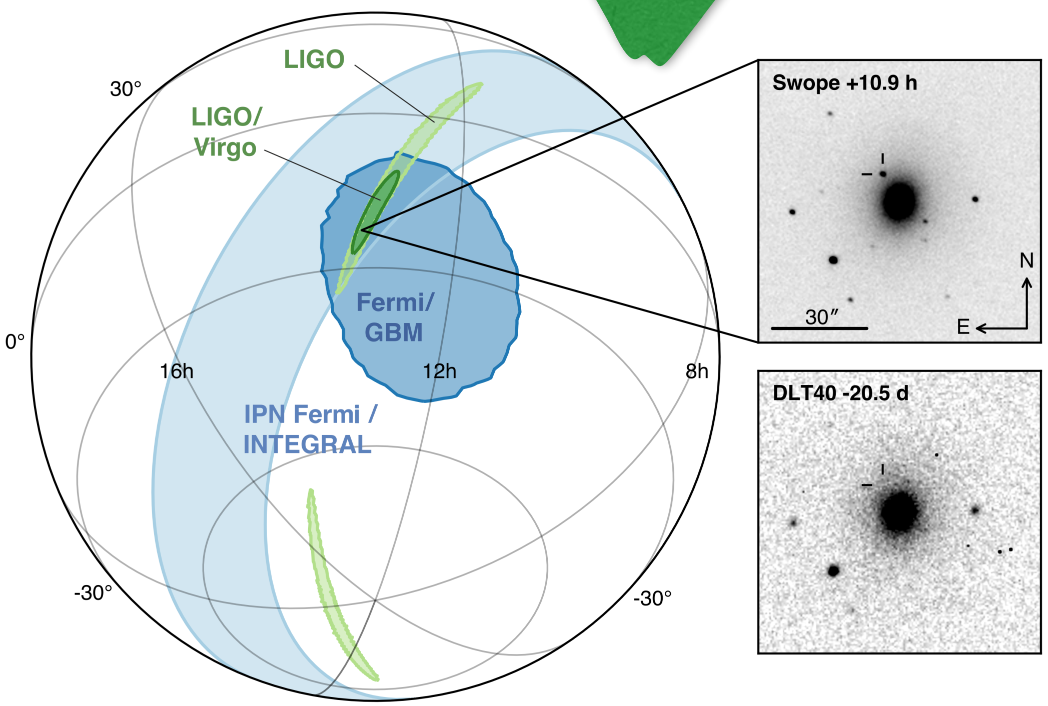


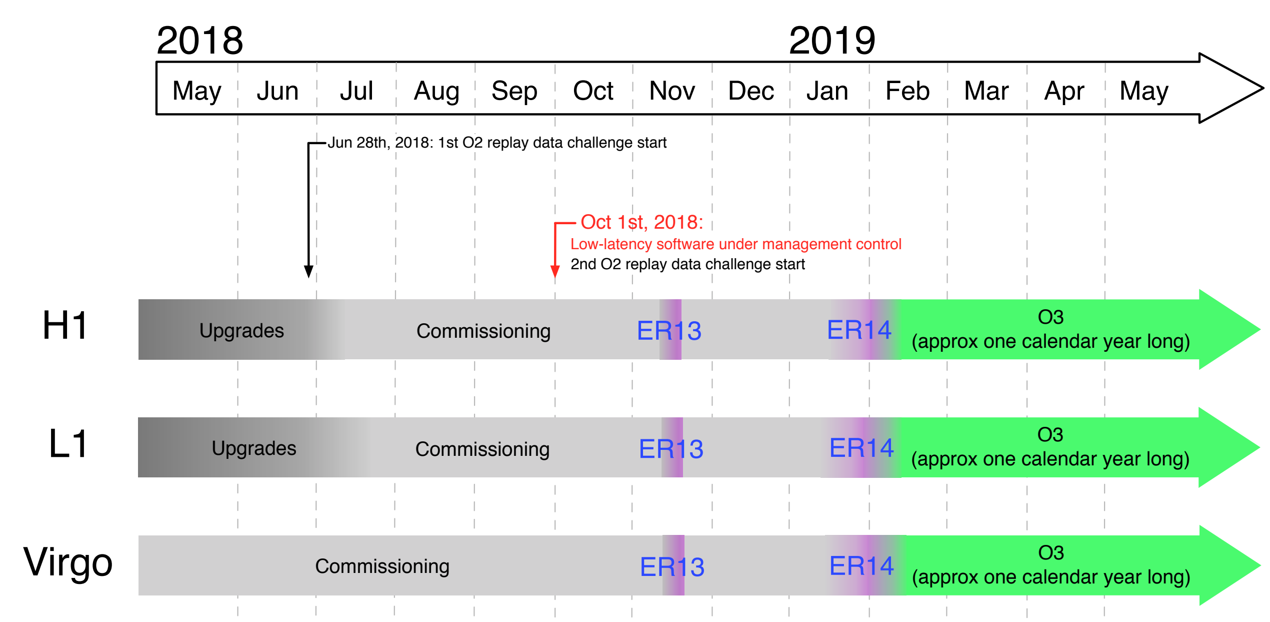
LSC VIRGO

University of Glasgow

Outlook

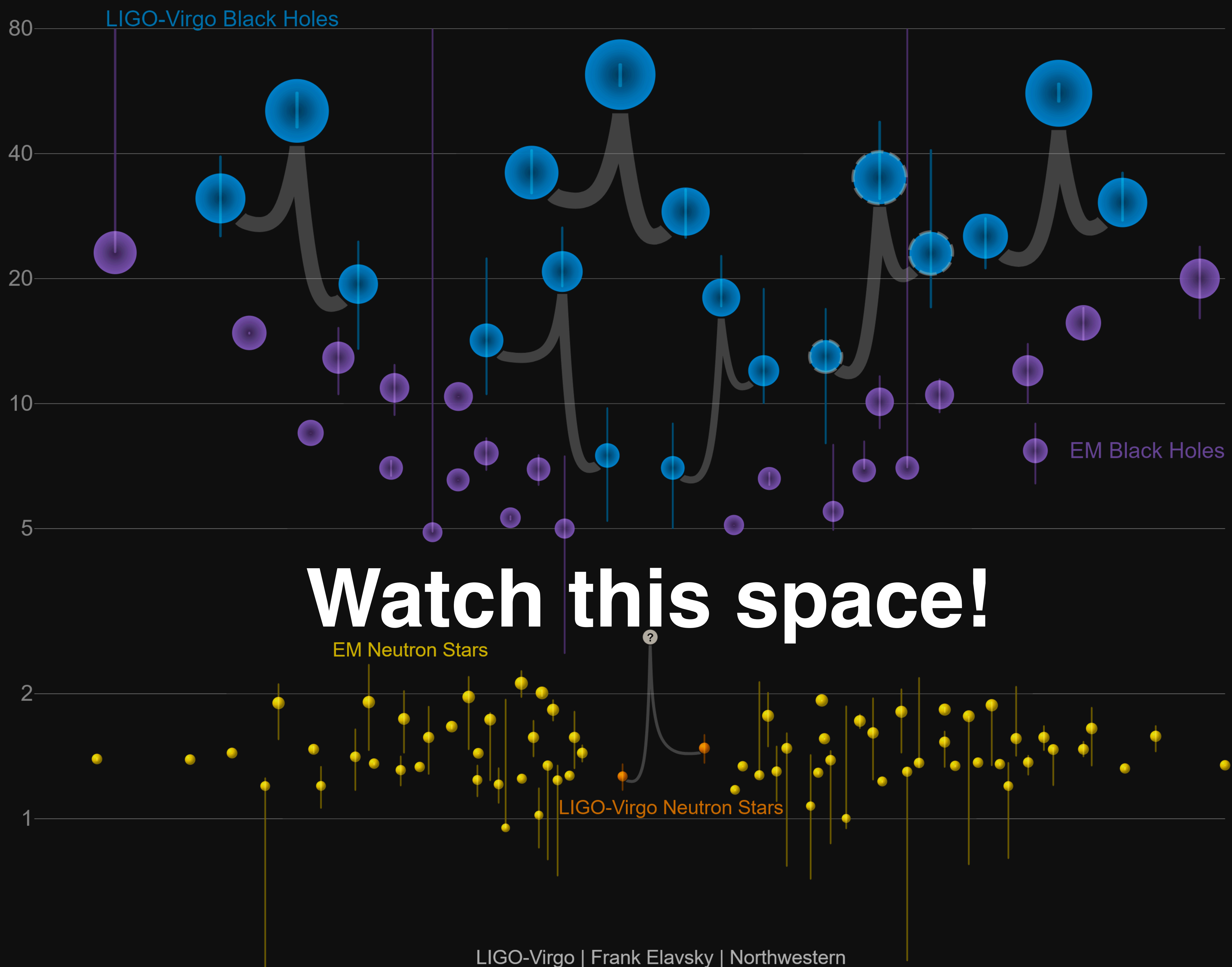
- Parameter estimation succeeded beautifully in characterising first BBH and BNS detections
- Spurred development of hierarchical models for population analyses
    - Understanding selection functions essential in GW astronomy
- O3 will deliver many more events:
    - ~1 BBH / week
    - BNS every month or two?
- Keeping up with data will be challenging!
    - Computational improvements in waveforms, inference algorithms
    - We are still analysing O2 data…
- Precision GW astronomy with populations will reveal systematics…





LSC VIRGO

University of Glasgow


LIGO-Virgo Black Holes
80
40
20
10
5
2
1
EM Black Holes
Watch this space!
?
EM Neutron Stars
LIGO-Virgo Neutron Stars
LIGO-Virgo | Frank Elavsky | Northwestern