Gravitational wave data analysis
Part II

Sylvain Marsat (L2IT, Toulouse)

Outline

Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

Part II

- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- PE results from LVK
- Future detectors and their challenges

Outline

Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

Part II

- **Bayesian parameter estimation basics, likelihood**
- Parameter space and waveforms
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- PE results from LVK
- Future detectors and their challenges

Bayes theorem and posterior distribution

Bayes theorem

Likelihood $p(d|\theta, M) = \mathcal{L}(d|\theta, M)$

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

Prior distribution
- a priori knowledge of parameters

Posterior distribution
- target of the analysis
- multidim. distribution, discrete samples

Evidence $p(d|M) = \int d\theta \, p(d|\theta, M)p(\theta|M)$
- normalization of the posterior
- important for model comparison

θ inferred params (17 for GW source)
d data (observed data in detector)
M model (context, assumptions)

The likelihood

Likelihood: $p(d|\theta)$

$$\mathcal{L} = p(\mathrm{data}|\mathrm{signal\ params})$$

Signal model: $h(\theta)$

- includes instrument response
- may be approximate

Assume calibrated data
(reality: marginalize over calibration)

$$d = h(\theta) + n$$
$$p(d|\theta) = p(n = d - h(\theta))$$

Likelihood: probability that the noise explains the residuals between data and signal

Whittle likelihood

For a stationary Gaussian process:
independence FD, diagonal covariance

$$\langle \tilde{n}_k \tilde{n}_l^* \rangle = \frac{1}{2\Delta f} S_n(f_k) \delta_{kl}$$
$$\mathrm{Re}\, \tilde{n}_k,\ \mathrm{Im}\, \tilde{n}_k \sim \mathcal{N}\left(0, \frac{1}{4\Delta f} S_n(f_k)\right)$$

The likelihood

Likelihood: $p(d|\theta)$

$$\mathcal{L} = p(\text{data}|\text{signal params})$$

Signal model: $h(\theta)$

- includes instrument response
- may be approximate

Assume calibrated data
(reality: marginalize over calibration)

$$d = h(\theta) + n$$

$$p(d|\theta) = p(n = d - h(\theta))$$

Likelihood: probability that the noise explains the residuals between data and signal

Whittle likelihood

For a stationary Gaussian process:
independence FD, diagonal covariance

$$\langle \tilde{n}_k \tilde{n}_l^* \rangle = \frac{1}{2\Delta f} S_n(f_k) \delta_{kl}$$

$$\mathrm{Re}\, \tilde{n}_k,\ \mathrm{Im}\, \tilde{n}_k \sim \mathcal{N}\left(0, \frac{1}{4\Delta f} S_n(f_k)\right)$$

$$\ln \mathcal{L} = \ln p(\mathbf{n}) = \sum_{k>0} \ln p(\tilde{n}_k)$$

The likelihood

Likelihood: $p(d|\theta)$

$$\mathcal{L} = p(\text{data}|\text{signal params})$$

Signal model: $h(\theta)$

- includes instrument response
- may be approximate

Assume calibrated data
(reality: marginalize over calibration)

$$d = h(\theta) + n$$
$$p(d|\theta) = p(n = d - h(\theta))$$

Likelihood: probability that the noise explains the residuals between data and signal

Whittle likelihood

For a stationary Gaussian process: independence FD, diagonal covariance

$$\langle \tilde{n}_k \tilde{n}_l^* \rangle = \frac{1}{2\Delta f} S_n(f_k) \delta_{kl}$$
$$\mathrm{Re}\, \tilde{n}_k,\ \mathrm{Im}\, \tilde{n}_k \sim \mathcal{N}\left(0, \frac{1}{4\Delta f} S_n(f_k)\right)$$

$$\begin{aligned}
\ln \mathcal{L} = \ln p(\mathbf{n}) &= \sum_{k>0} \ln p(\tilde{n}_k) \\
&= \sum_{k>0} -\frac{1}{2} \frac{4\Delta f}{S_n(f_k)} |\tilde{n}_k|^2 + \cancel{\text{const}} \\
&= -\frac{1}{2} 4 \int_{f>0} \frac{df}{S_n(f)} |\tilde{n}(f)|^2 \\
&= -\frac{1}{2} (n|n)
\end{aligned}$$

The likelihood

Likelihood: $p(d|\theta)$

$$\mathcal{L} = p(\text{data}|\text{signal params})$$

Signal model: $h(\theta)$

- includes instrument response
- may be approximate

Assume calibrated data
(reality: marginalize over calibration)

$$d = h(\theta) + n$$

$$p(d|\theta) = p(n = d - h(\theta))$$

Likelihood: probability that the noise explains the residuals between data and signal

Whittle likelihood

For a stationary Gaussian process: independence FD, diagonal covariance

$$\langle \tilde{n}_k \tilde{n}_l^* \rangle = \frac{1}{2\Delta f} S_n(f_k) \delta_{kl}$$

$$\mathrm{Re}\, \tilde{n}_k,\ \mathrm{Im}\, \tilde{n}_k \sim \mathcal{N}\left(0, \frac{1}{4\Delta f} S_n(f_k)\right)$$

$$\begin{aligned}
\ln \mathcal{L} = \ln p(\mathbf{n}) &= \sum_{k>0} \ln p(\tilde{n}_k) \\
&= \sum_{k>0} -\frac{1}{2} \frac{4\Delta f}{S_n(f_k)} |\tilde{n}_k|^2 + \cancel{\text{const}} \\
&= -\frac{1}{2} 4 \int_{f>0} \frac{df}{S_n(f)} |\tilde{n}(f)|^2 \\
&= -\frac{1}{2} (n|n)
\end{aligned}$$

$$\boxed{\ln \mathcal{L}(\theta) = -\frac{1}{2} (h(\theta) - d | h(\theta) - d)}$$

Norm of residuals !

TD/FD covariance

Limitations to Whittle:

Covariance matrix for noise vector in time domain: Toeplitz structure

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

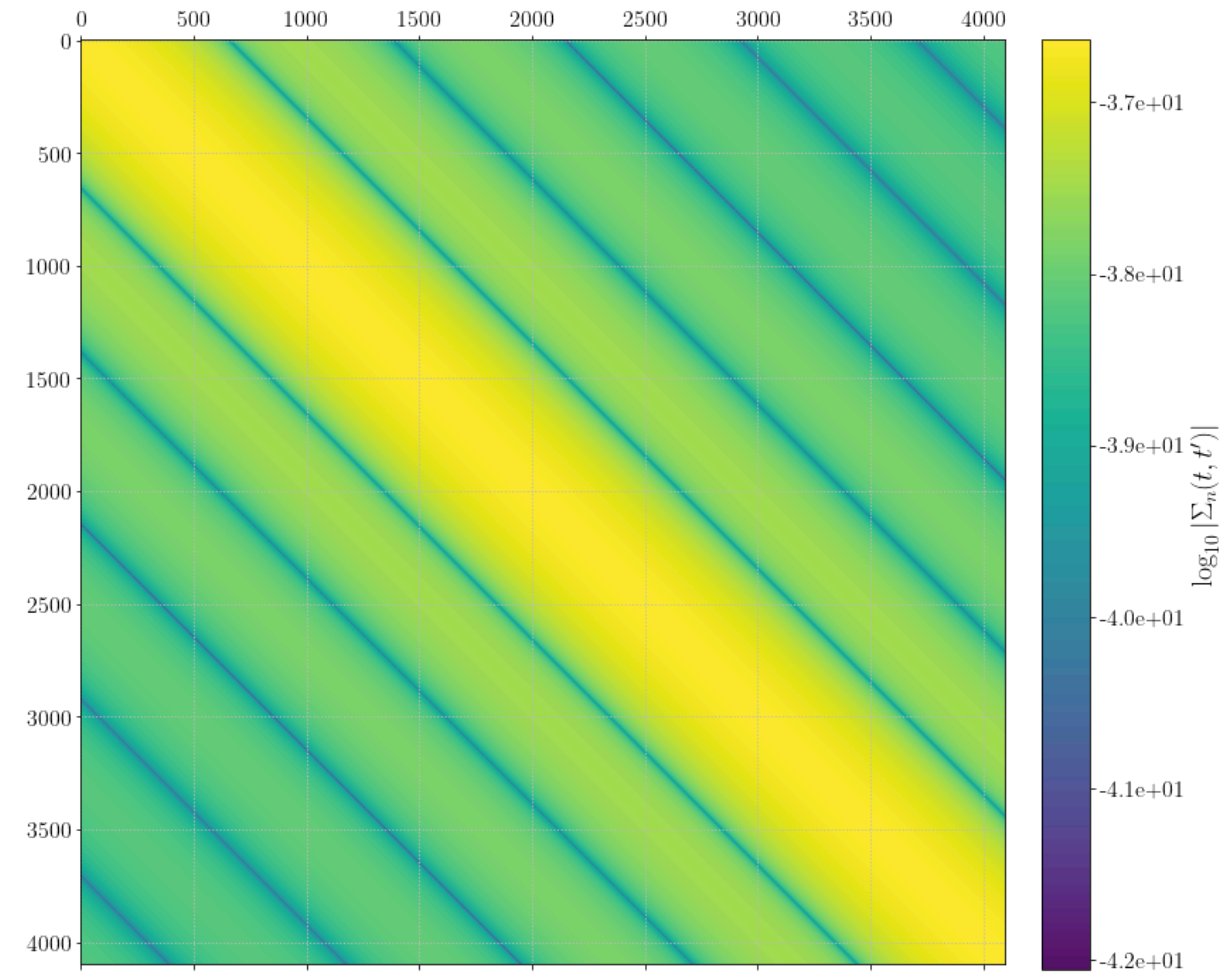
$$C(t,t') \equiv C(t-t')$$

Diagonality after DFT requires in fact Circulant structure (periodicity)

$$C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{bmatrix}$$

Time domain covariance

Can work directly in time domain with Toeplitz structure



- Ringdown analysis: selection of post-merger times
- Dealing with data gaps

The evidence and Bayes ratio

Evidence

$$p(d|M) = \int d\theta \, p(d|\theta, M) p(\theta|M)$$

- ignored in parameter estimation, normalization constant
- multidimensional integral, hard to compute

Bayes factor

$$B_{12} = \frac{p(d|M_1)}{p(d|M_2)}$$

- model comparison (usually log)
- penalty for overfitting

$$\frac{p(M_1|d)}{p(M_2|d)} = \frac{p(d|M_1)}{p(d|M_2)} \frac{p(M_1)}{p(M_2)}$$

[Kass-Raftery 1995]

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

Outline

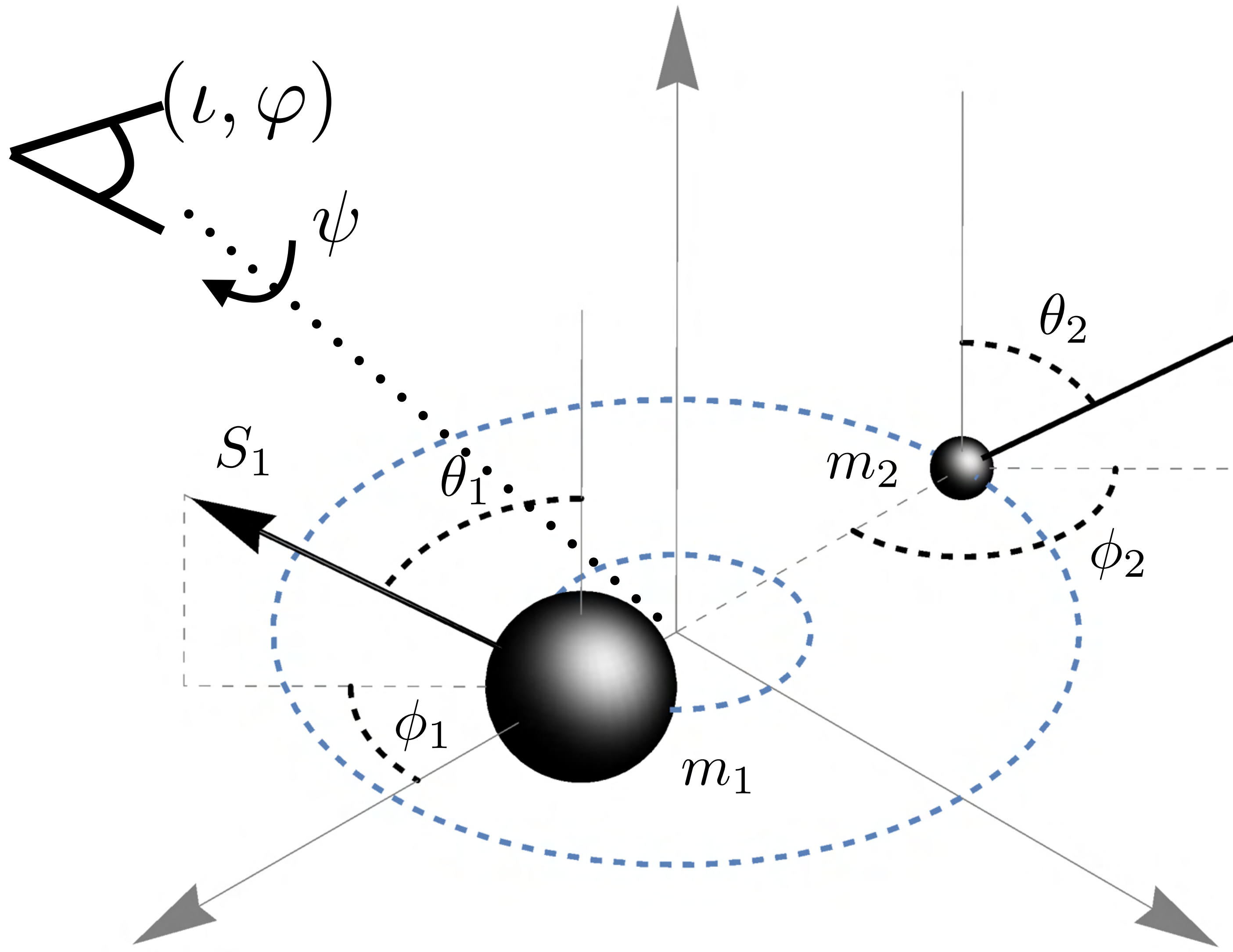
Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

Part II

- Bayesian parameter estimation basics, likelihood
- **Parameter space and waveforms**
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- PE results from LVK
- Future detectors and their challenges

The parameter space and priors



CBC: 15+2+2 parameters

- intrinsic: 2 masses, 2*3 spin vectors
- distance: 1
- time of coalescence: 1
- direction to the observer: 2 angles
- sky position in observer's frame: 2 angles
- polarization angle: 1 angle
- +eccentricity, periastron: 2
- +tidal deformabilities BNS: 2

Intrinsic parameters

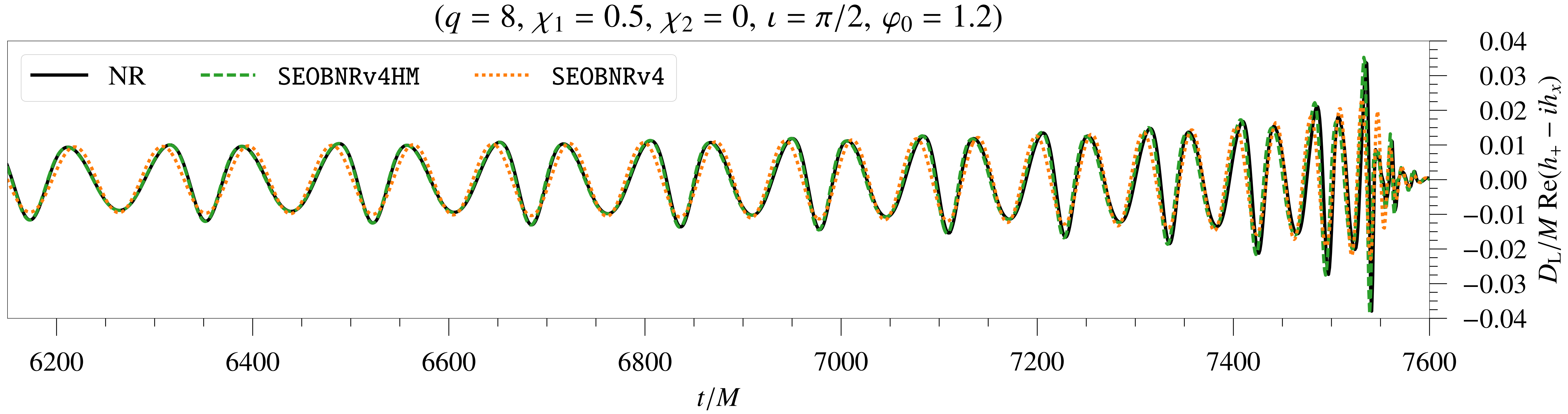
- masses, spins, tidal deformabilities, eccentricity
- expensive: generate GR solution
- priors: physically motivated, but arbitrary

Extrinsic parameters

- distance, time, orientation angles
- cheap: simple geometry of source/ detector
- prior for distance: uniform in volume ?
- prior for time: uniform
- prior for angles: uniform (on a sphere)

Waveform complexity I

Higher harmonics beyond h22: $$h_+ - ih_\times = \sum_{\ell\geq 2}\sum_{m=-\ell}^{+\ell} {}_{-2}Y_{\ell m}(\iota,\varphi)h_{\ell m} \qquad h_{\ell m} \propto e^{-im\phi_{\mathrm{orb}}}$$

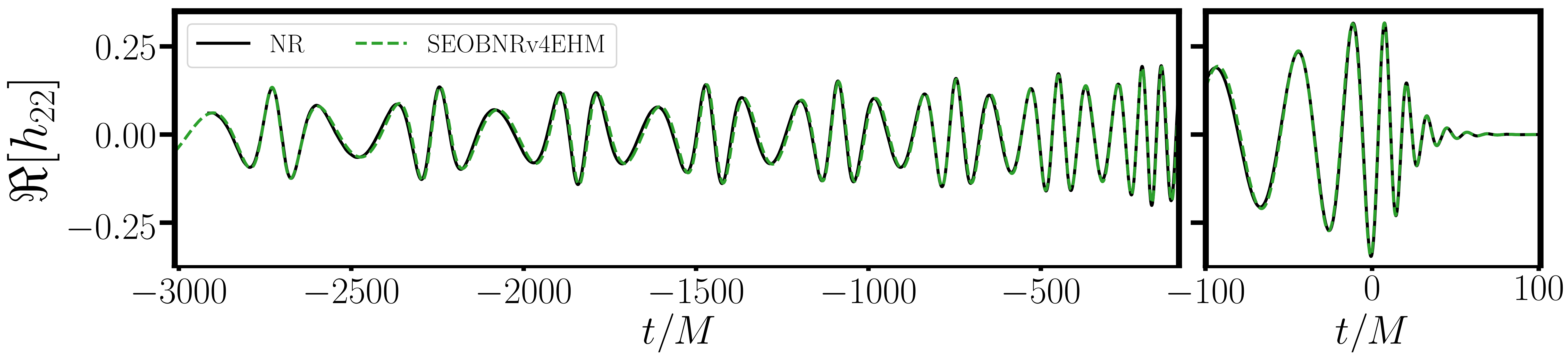


- stronger for high inclination, high mass ratio
- most important at merger

Eccentricity:

- creates another set of harmonics
- fast circularization before merger

[arXiv:2112.06952]



Waveform complexity II

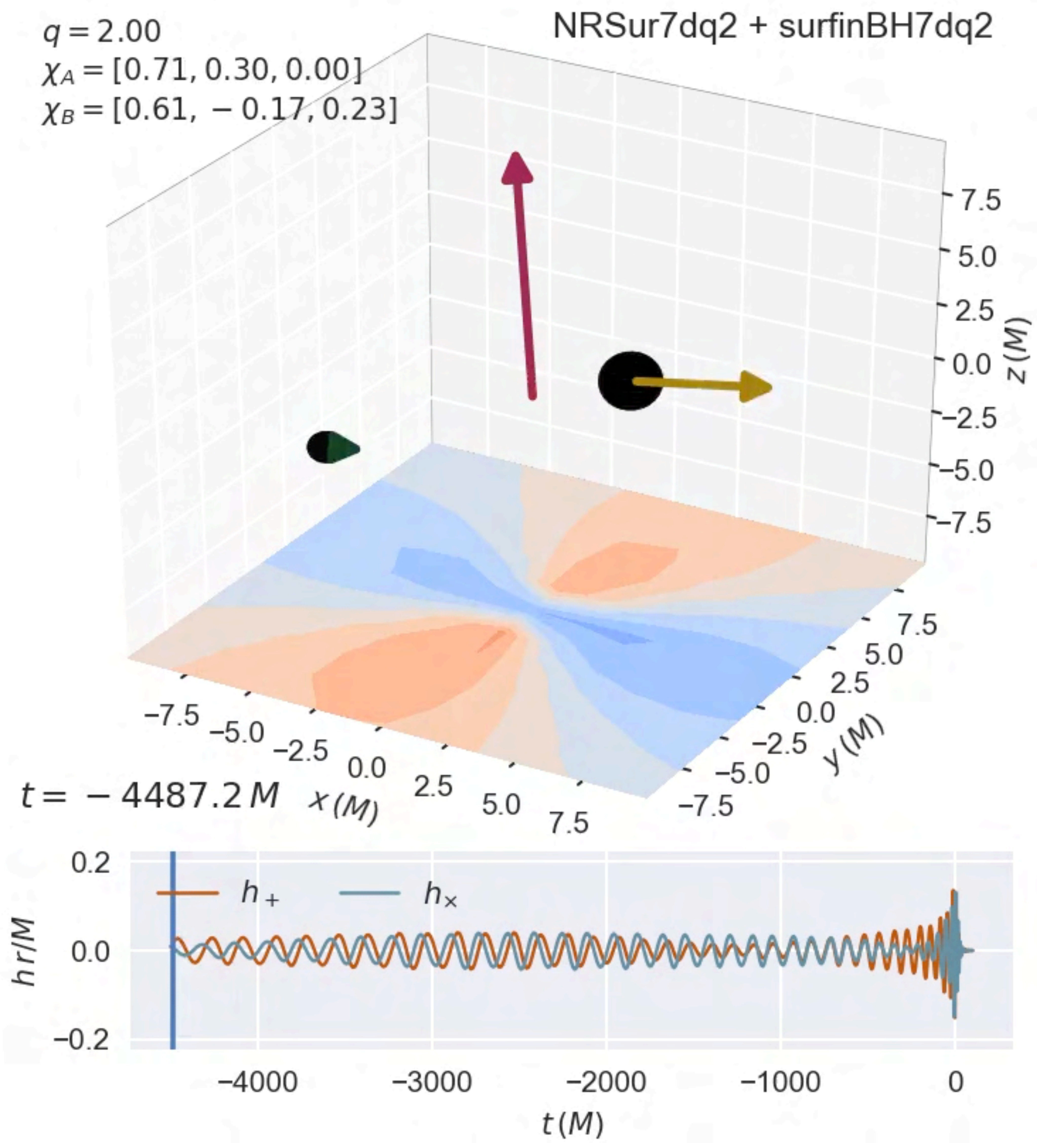
Effect of aligned spins

- Aligned/anti-aligned spins: longer/shorter inspiral, reaching higher/lower frequencies
- Degeneracy with mass ratio

Effect of precession

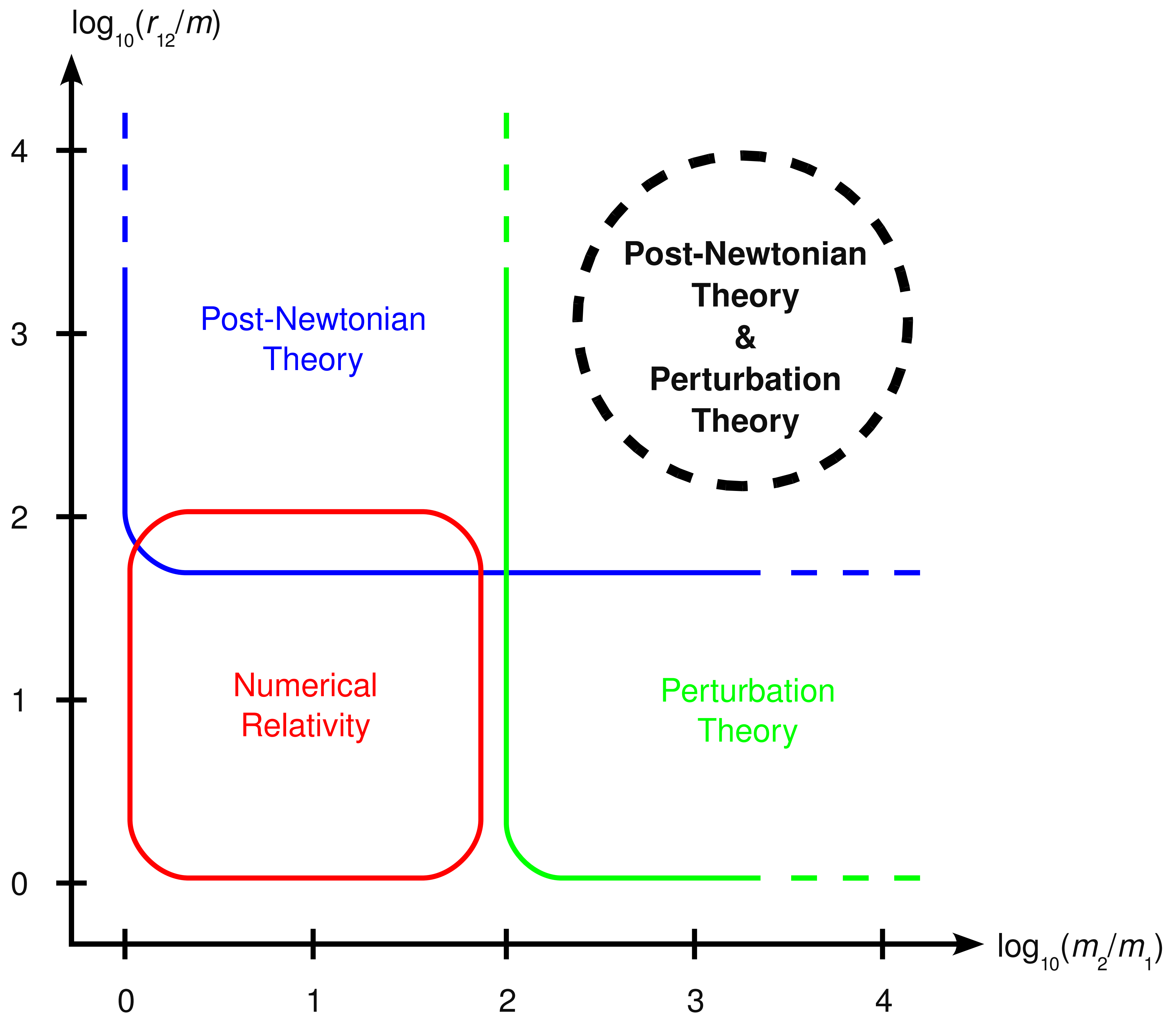
- introduce Precessing frame, follows plane of the orbit
- time-dependent rotation to Inertial frame
- not exact, mode asymmetries !

$$h^{\mathrm{P}}_{\ell m} = \sum_{m=-\ell}^{\ell} \mathcal{D}^{\ell}_{m' m}(\alpha, \beta, \gamma) h^{\mathrm{I}}_{\ell m'}$$



[https://vijayvarma392.github.io/binaryBHexp/]

Waveform models



Analytical/numerical methods

- post-Newtonian, post-Minkowskian
- Gravitational Self-Force
- Numerical Relativity

Computational approaches

We need millions of waveform evaluations !

- Phenomenological waveforms: analytic fits
- EOB waveforms: integrate ODE
- NR: costly large-scale simulations

Acceleration of EOB, NR with Reduced Order Models (ROMs), surrogates

Acceleration of likelihoods (heterodyning, …)

Frontiers of waveform modelling

- eccentric IMR waveforms
- EMRIs: Extreme (Intermediate) Mass Ratio Inspirals
- matter effects for BNS, BNS merger
- environmental effects

Outline

Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

Part II

- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- **Fisher matrix approach**
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- PE results from LVK
- Future detectors and their challenges

Fisher matrix approach

Fisher matrix and covariance

Objective: simple simulation of PE

Local Taylor expansion of likelihood around true parameters, ignore noise:

$$\ln \mathcal{L} = -\frac{1}{2}(h(\theta) - h(\theta_0)|h(\theta) - h(\theta_0))$$

$$h(\theta) = h(\theta_0) + \Delta\theta_i \partial_i h + \mathcal{O}(\Delta\theta^2)$$

$$\ln \mathcal{L} = -\frac{1}{2} F_{ij} \Delta\theta_i \Delta\theta_j$$

$F_{ij} \equiv (\partial_i h|\partial_j h)$ Fisher matrix

$\mathbf{\Sigma} = \mathbf{F}^{-1}$ Fisher covariance

Gaussian approximation of the posterior, with Fisher covariance

Parameter biases

In the presence of a residual δh (noise, error in the waveform): bias in best-fit parameters

$$\Delta\theta_i = F^{-1}_{ij}(\partial_j h|\delta h)$$

In practice...

Valid in the high-SNR limit

- requires signal derivatives
- sensitive to degeneracies (even at high SNR)
- numerically delicate !

Effect of noise on posterior

Effect of noise $d = h_0 + n$

$$\ln \mathcal{L} = \underbrace{-\frac{1}{2}(h - h_0|h - h_0)}_{\text{0-noise likelihood}} + \underbrace{(h - h_0|n)}_{\text{cross-term}} \underbrace{- \frac{1}{2}(n|n)}_{\text{const.}}$$

0-noise likelihood

cross-term

changes with n

const.

Effect of noise on posterior

Effect of noise $d = h_0 + n$

$$\ln \mathcal{L} = -\frac{1}{2}(h - h_0|h - h_0) + (h - h_0|n) - \frac{1}{2}(n|n)$$

0-noise likelihood

cross-term

changes with n

const.

Bias in Fisher approach:

$$\delta h = n \qquad \Delta\theta_i = F_{ij}^{-1}(\partial_j h|n)$$

$$\langle \Delta\theta_i \Delta\theta_j \rangle = F_{ik}^{-1} F_{jl}^{-1} \langle (\partial_i h|n)(n|\partial_j h) \rangle$$

$$\langle |n)(n| \rangle = \mathbf{1}$$

$$\langle \Delta\theta_i \Delta\theta_j \rangle = F_{ij}^{-1} = \Sigma_{ij}$$

Biases due to noise follow the Fisher covariance

Effect of noise on posterior

Effect of noise $d = h_0 + n$

$$\ln \mathcal{L} = -\frac{1}{2}(h - h_0|h - h_0) + (h - h_0|n) - \frac{1}{2}(n|n)$$

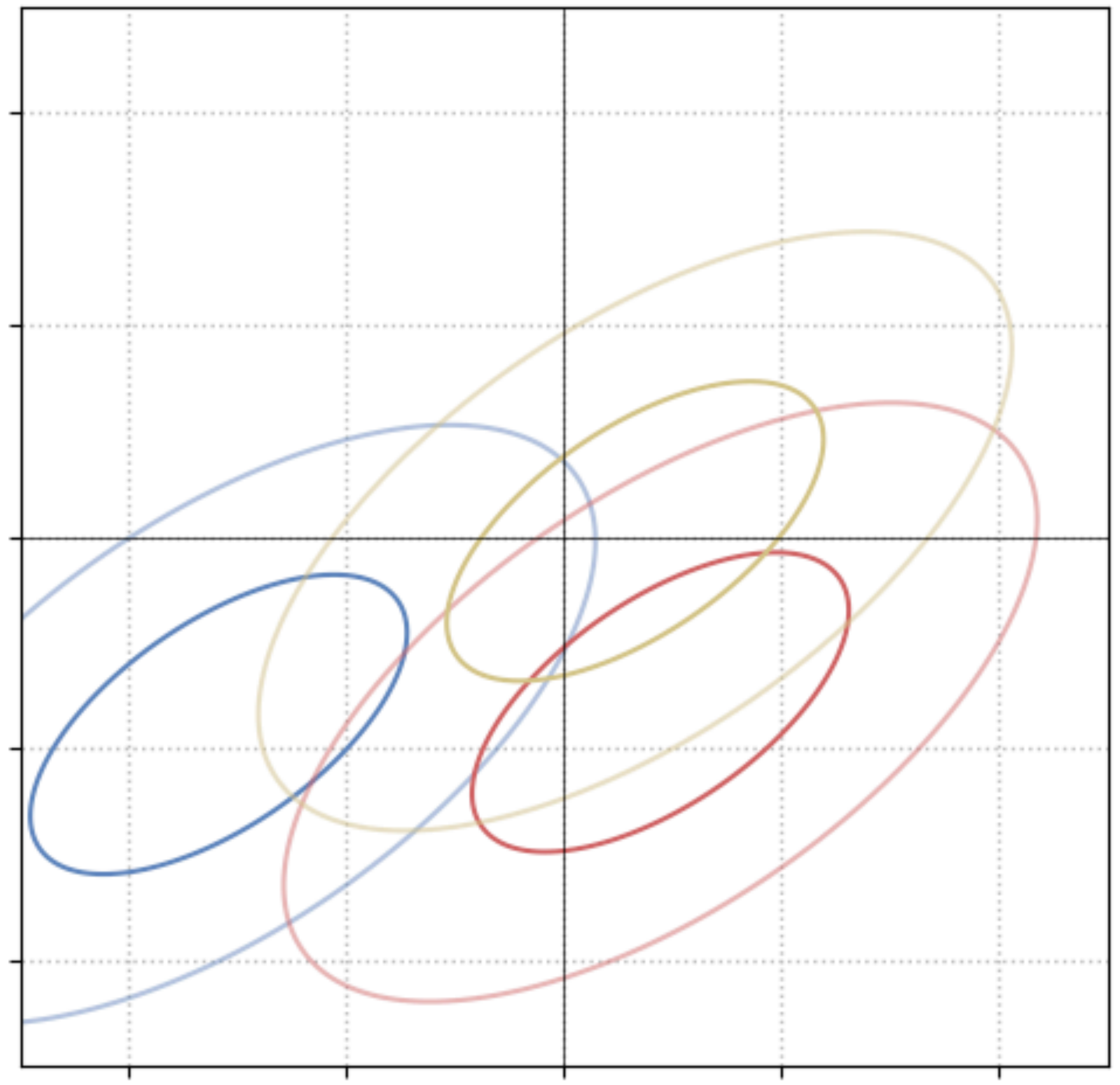
0-noise likelihood

cross-term
changes with n

const.

For different noise realizations:

- peak of the posterior moves around by the size of the statistical errors
- width of posterior unaffected (in this approx.)



Bias in Fisher approach:

$$\delta h = n \qquad \Delta\theta_i = F^{-1}_{ij}(\partial_j h|n)$$

$$\langle \Delta\theta_i \Delta\theta_j \rangle = F^{-1}_{ik} F^{-1}_{jl} \langle (\partial_i h|n)(n|\partial_j h) \rangle$$

$$\langle |n)(n| \rangle = \mathbf{1}$$

$$\langle \Delta\theta_i \Delta\theta_j \rangle = F^{-1}_{ij} = \Sigma_{ij}$$

Biases due to noise follow the Fisher covariance

Outline

Part I

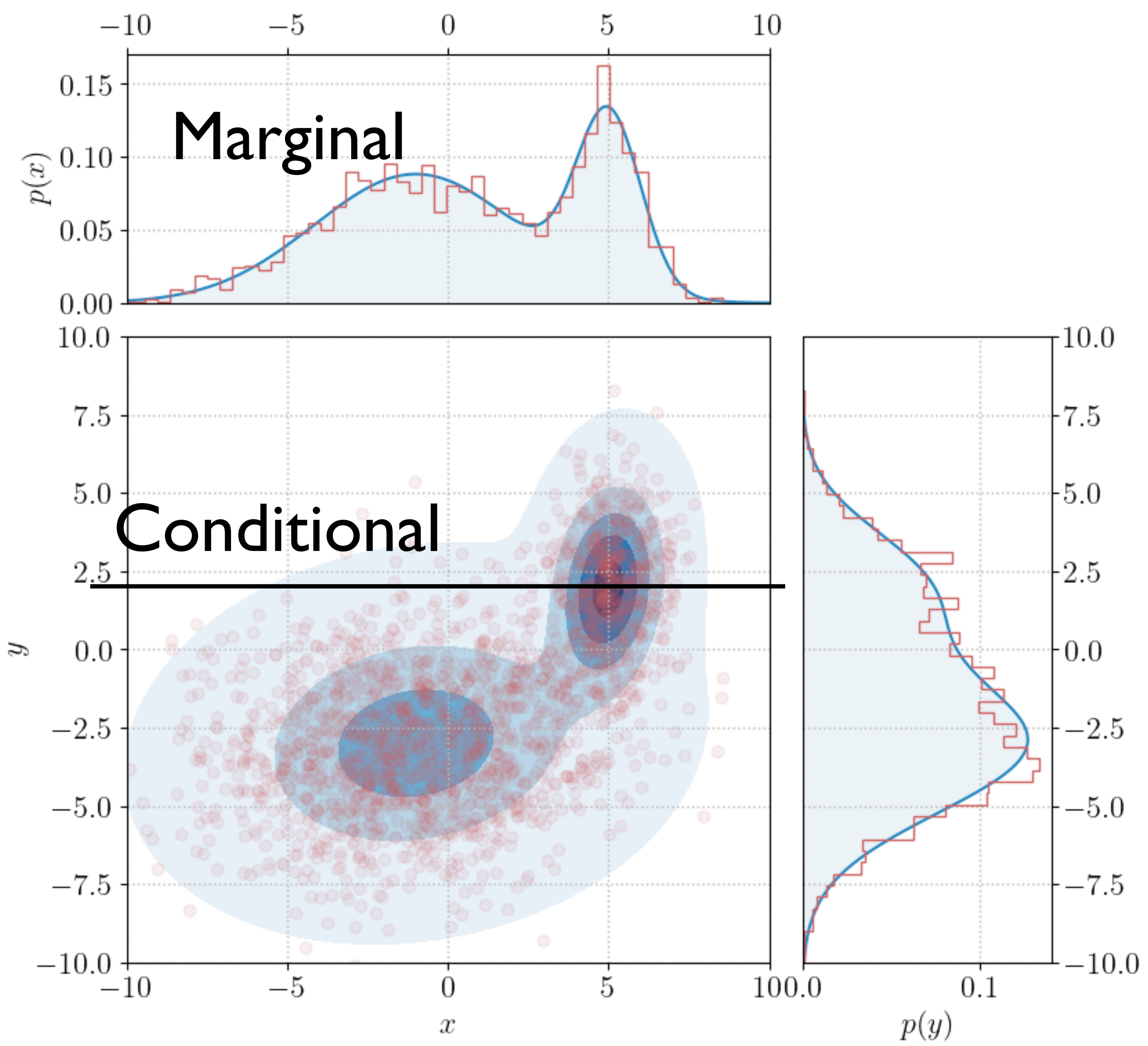
- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

Part II

- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- Fisher matrix approach
- **Metropolis-Hastings MCMC, Parallel tempering and example PE**
- PE toolbox
- PE results from LVK
- Future detectors and their challenges

Sampling: introduction

Multidimensional posteriors



Marginal:

$$p(x) = \int dy\, p(x,y)$$

Conditional:

$$p(x|y)$$

Curse of dimensionality

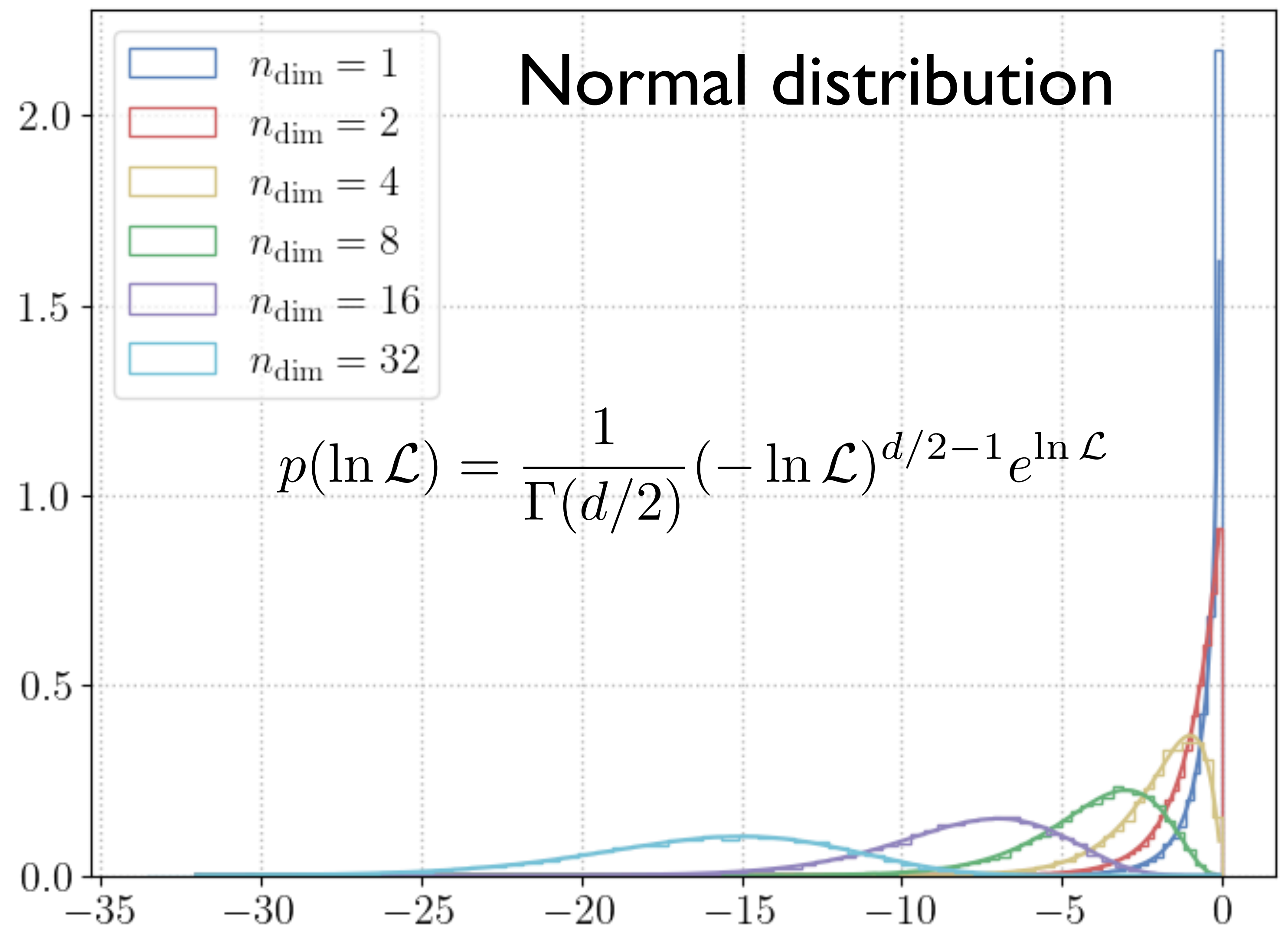
- Grids explode:

$$N^d$$

- Relevant volume vs full volume:

$$v/V = (\ell/L)^d$$

- In high dimensions, tails are important



Posterior samples

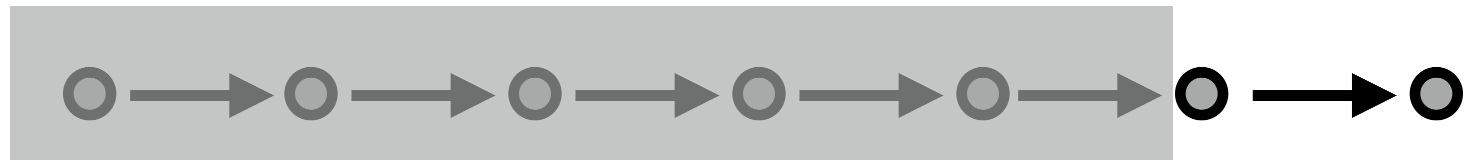
Samples: independent draws from the pdf

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) \sim_{n\to+\infty} \int dx\, p(x) f(x)$$

- Mandatory in high dimensions !
- Convenient compression
- Change of variable trivial
- Marginalization trivial

MH MCMC

Markov Chain Monte Carlo



- chain of values, no memory
- jump proposal, proba. acceptance

MCMC for sampling

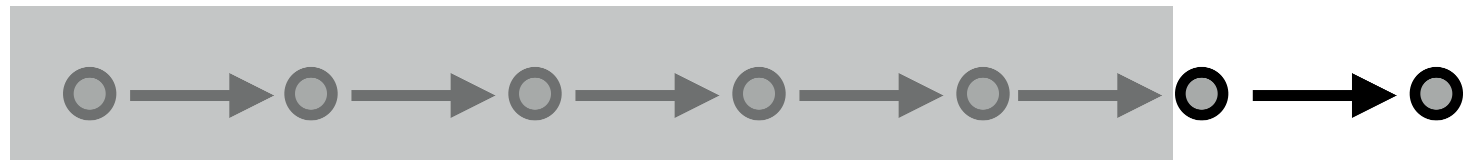
- ergodicity (hard !)
- stationarity of the distribution

$p(x)$ target prob. distribution

$T(x, y)$ transition prob. $x \to y$

MH MCMC

Markov Chain Monte Carlo



- chain of values, no memory
- jump proposal, proba. acceptance

MCMC for sampling

- ergodicity (hard !)
- stationarity of the distribution

$p(x)$ target prob. distribution

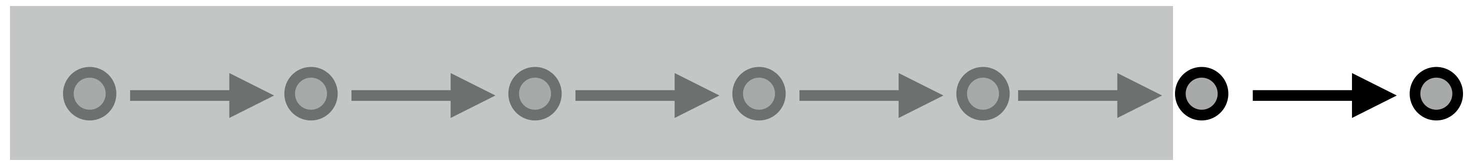
$T(x,y)$ transition prob. $x \to y$

Stationarity: $p(y) = \int dx\, p(x) T(x,y)$

Detailed balance
(sufficient): $p(x)T(x,y) = p(y)T(y,x)$

MH MCMC

Markov Chain Monte Carlo



- chain of values, no memory
- jump proposal, proba. acceptance

MCMC for sampling

- ergodicity (hard !)
- stationarity of the distribution

$p(x)$ target prob. distribution

$T(x,y)$ transition prob. $x \to y$

Stationarity: $p(y) = \int dx\, p(x) T(x,y)$

Detailed balance
(sufficient): $p(x)T(x,y) = p(y)T(y,x)$

Metropolis-Hastings MCMC

For generic proposal, build acceptance probability that respects detailed balance

$q(x,y)$ jump proposal $x \to y$

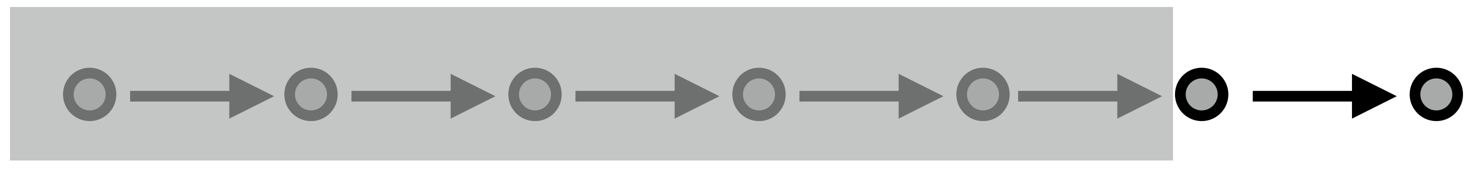
$\alpha(x,y)$ acceptance probability

$T(x,y) = q(x,y)\alpha(x,y)$

Reject jump with probability $1 - \alpha(x,y)$

MH MCMC

Markov Chain Monte Carlo



- chain of values, no memory
- jump proposal, proba. acceptance

MCMC for sampling

- ergodicity (hard !)
- stationarity of the distribution

$p(x)$ target prob. distribution

$T(x,y)$ transition prob. $x \to y$

Stationarity: $p(y) = \int dx\, p(x) T(x,y)$

Detailed balance (sufficient): $p(x)T(x,y) = p(y)T(y,x)$

Metropolis-Hastings MCMC

For generic proposal, build acceptance probability that respects detailed balance

$q(x,y)$ jump proposal $x \to y$

$\alpha(x,y)$ acceptance probability

$T(x,y) = q(x,y)\alpha(x,y)$

Reject jump with probability $1 - \alpha(x,y)$

MH: $\alpha(x,y) = \min\left(1, \frac{p(y)}{p(x)} \frac{q(y,x)}{q(x,y)}\right)$

Symmetric proposal: $\alpha(x,y) = \min\left(1, \frac{p(y)}{p(x)}\right)$

- going up: always accept !
- going down: accept or reject

MCMC proposals

Tradeoff:
- narrow: good acceptance, bad exploration
- wide: bad acceptance, good exploration

Simple proposals

Example: Gaussian proposal

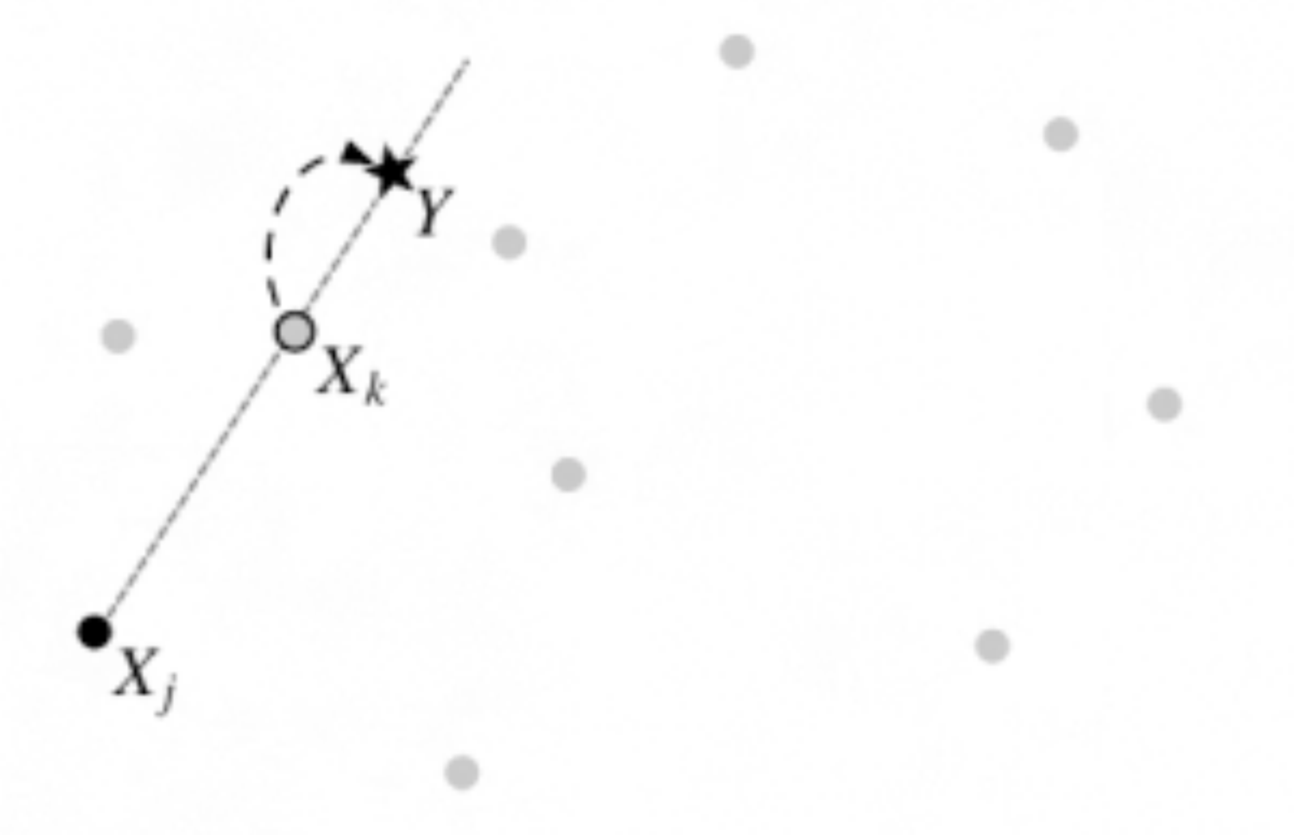
Covariance from Fisher ?
Adaptive covariance from chain ?

Tailored proposals

If multimodalities are known, propose non-local jumps to other modes

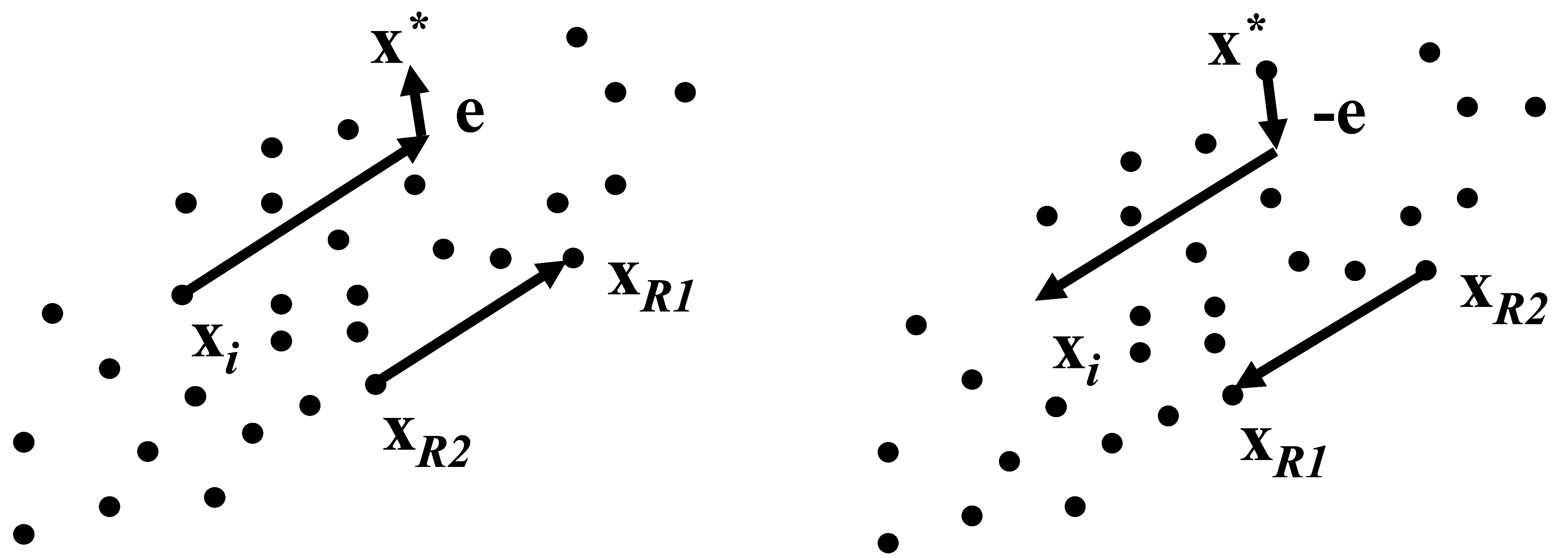
Ensemble sampling

Evolve chains in parallel, that will use each other to build proposal
Affine-invariant



Dynamical evolution

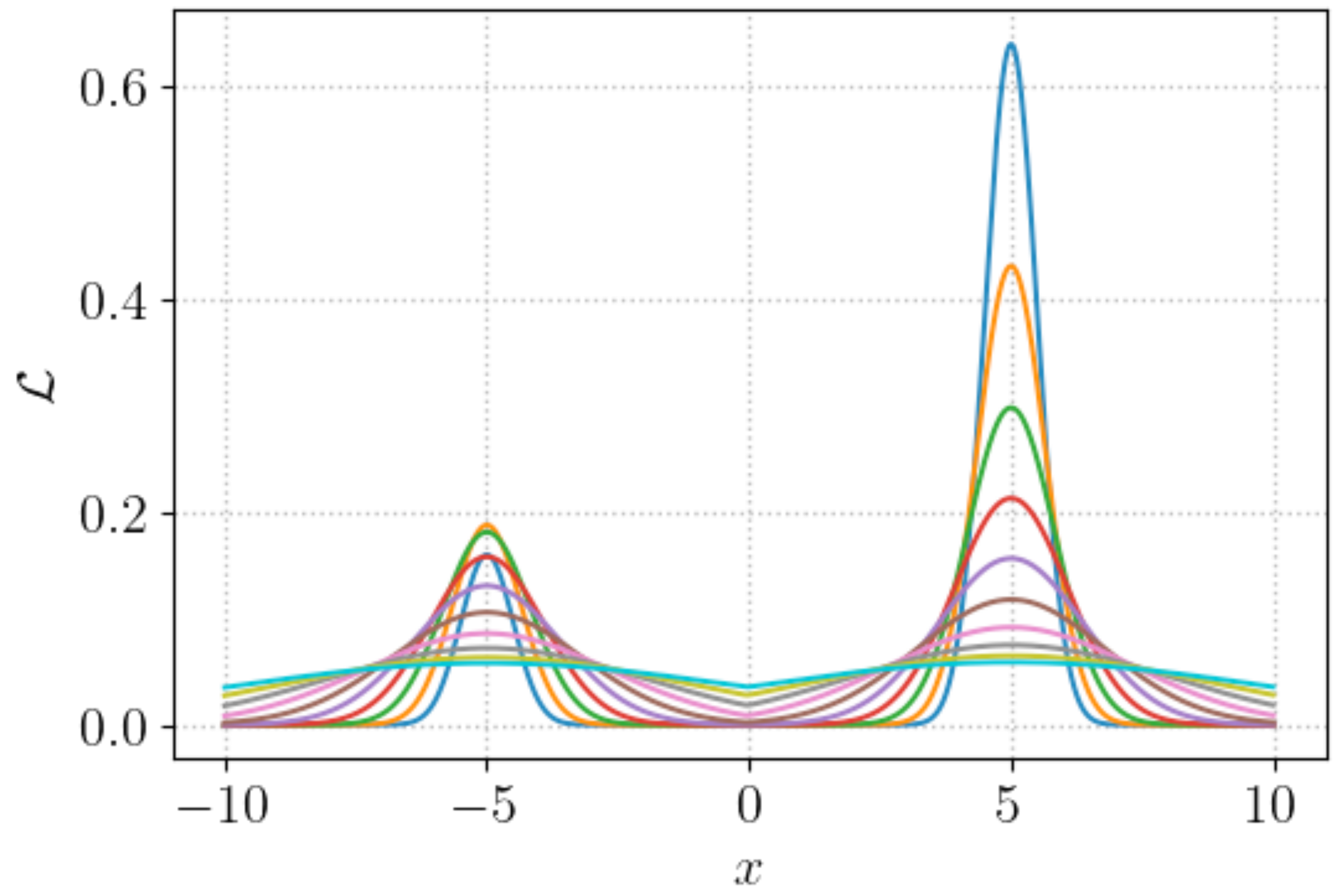
Use past of the chain to build proposal

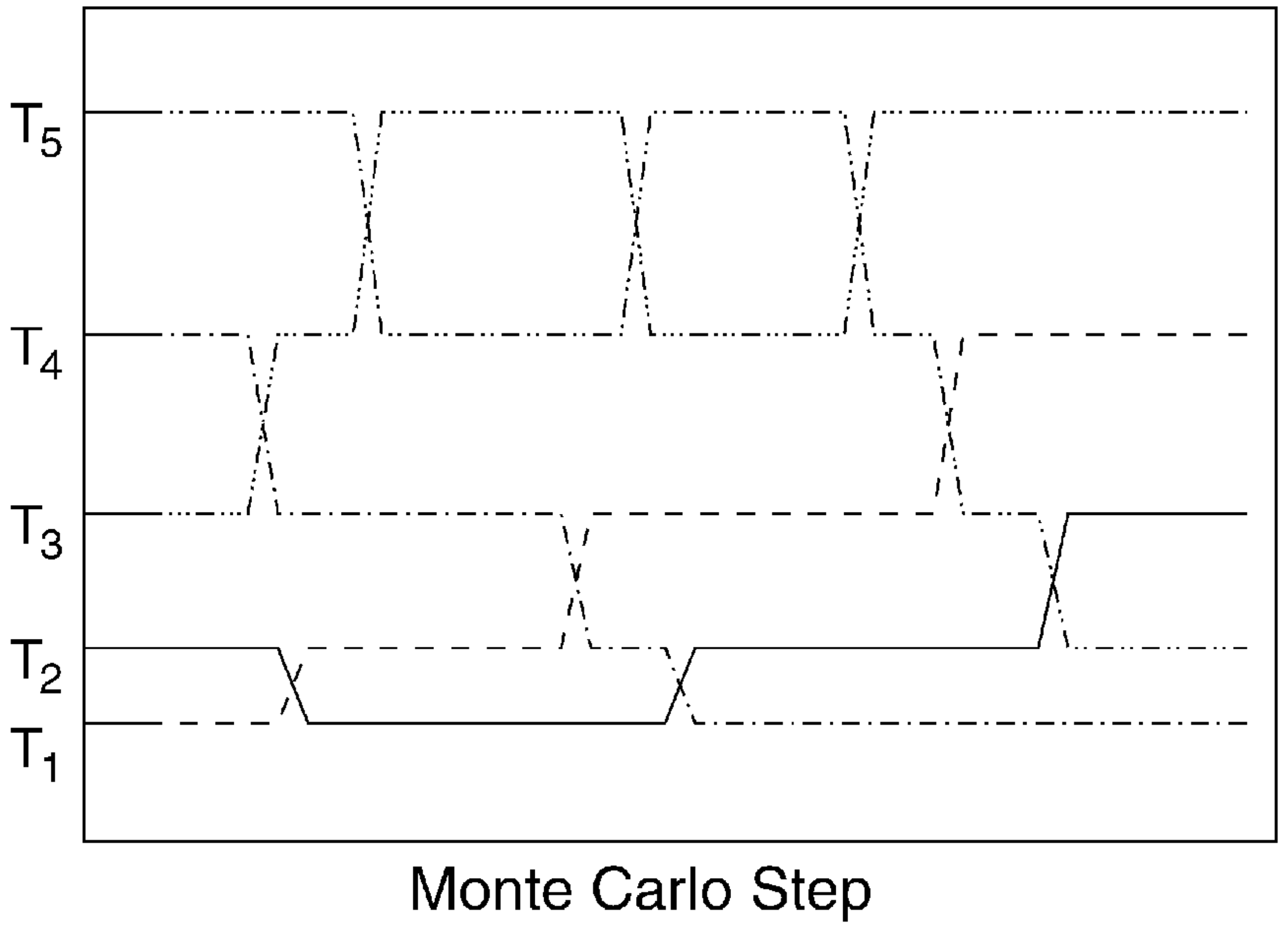


Multimodality issue and idea

Multimodality can require very long time to explore !

Idea: explore flattened (tempered) likelihood surfaces with other chains, exchange information

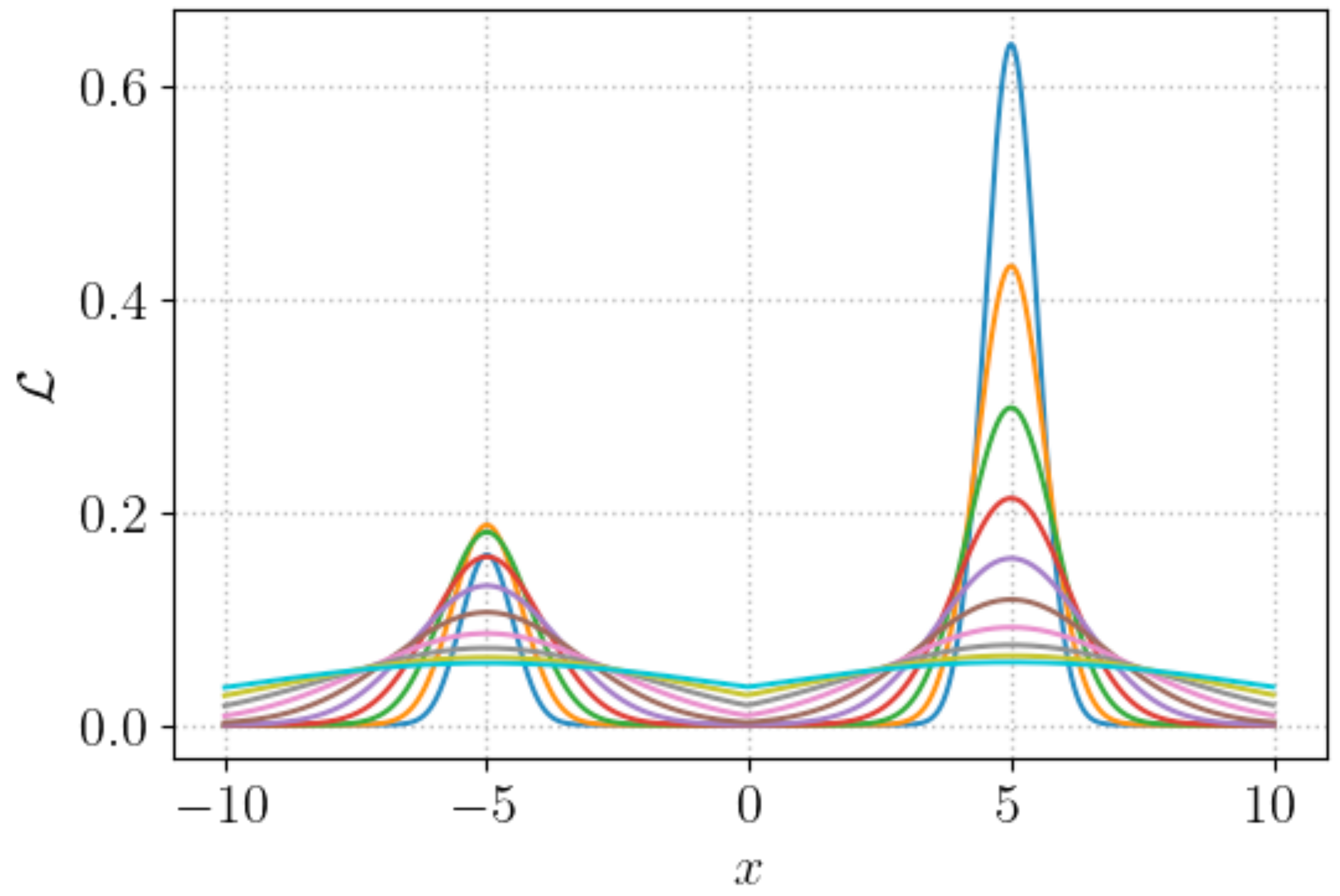


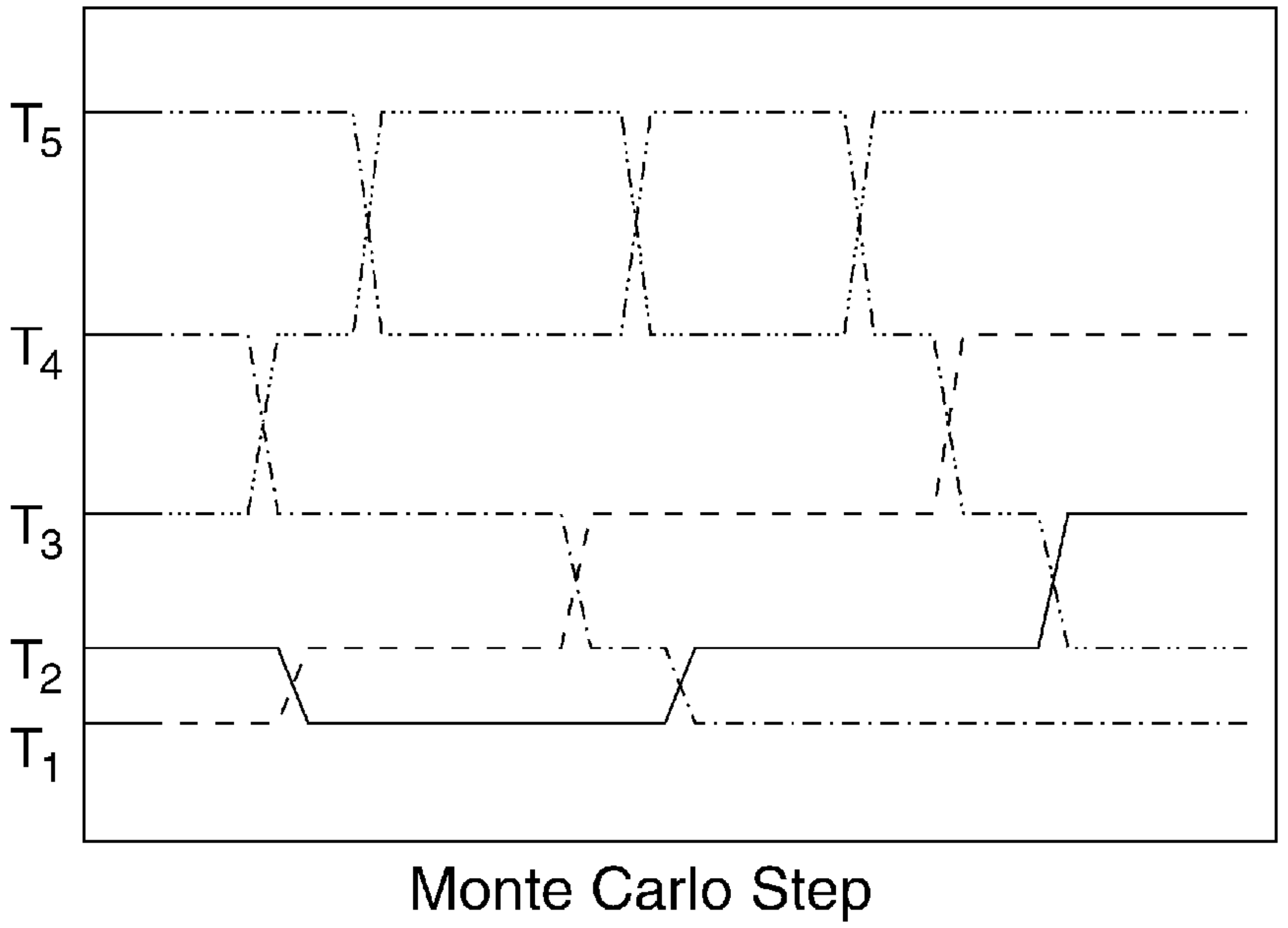


Multimodality issue and idea

Multimodality can require very long time to explore !

Idea: explore flattened (tempered) likelihood surfaces with other chains, exchange information





Replica exchange

Independent replicas:

$$p(x_1, \ldots, x_K) = \prod_{k=1}^{K} p_k(x_k)$$

Propose swap as MH step: ex.

$$(x_1, x_2, \ldots) \to (x_2, x_1, \ldots)$$

$$\alpha = \min\left(1, \frac{p_1(x_2)p_2(x_1)}{p_1(x_1)p_2(x_2)}\right)$$

Parallel tempering

prior

Tempered likelihoods: $p_k(x_k) = \pi(x_k)\mathcal{L}(x_k)^{\beta_k}$

$$\beta_k = 1/T_k$$

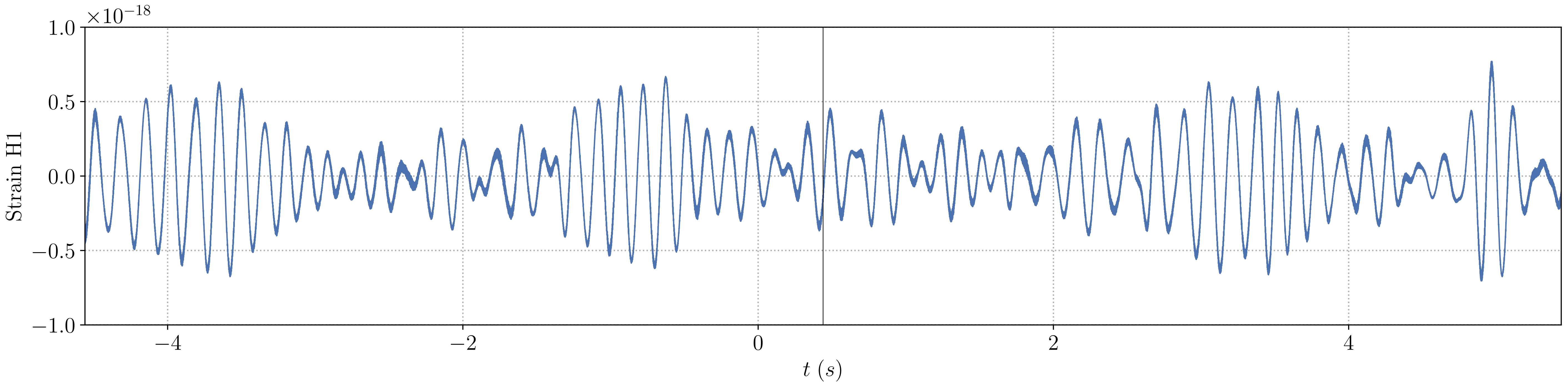
Swap acceptance:

$$\alpha_{ij} = \min\left(1, \left(\frac{\mathcal{L}(x_j)}{\mathcal{L}(x_i)}\right)^{\beta_i - \beta_j}\right)$$

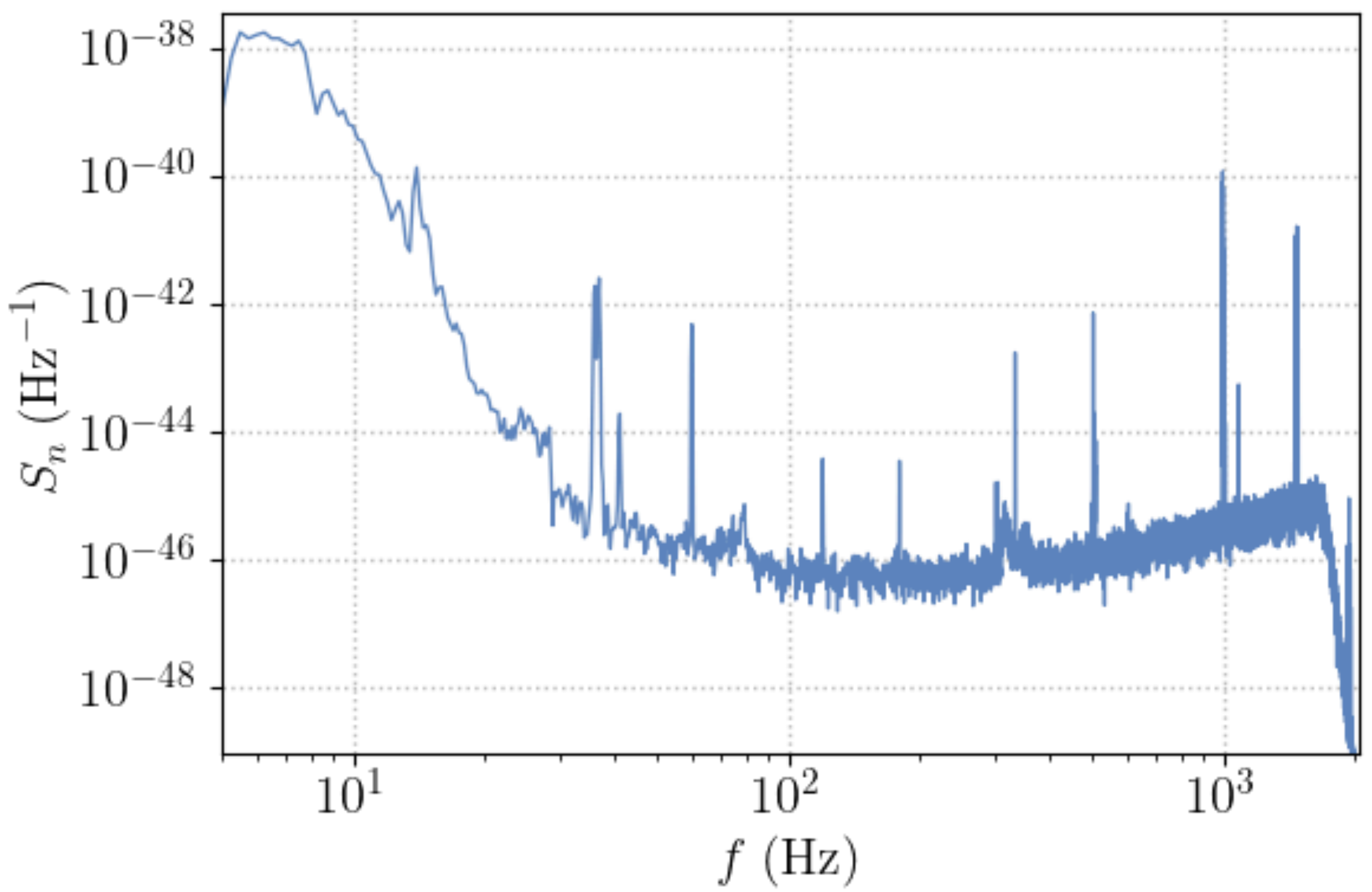
Adaptive temperatures to improve swaps

MH MCMC example: setting

Data from GWOSC



Estimated PSD (Welch)



- Simplified waveform model: from PhenomD

$$\tilde{h}(f) = A e^{i\alpha} e^{2i\pi f \Delta t} \tilde{h}_{22}^{\mathrm{PhenomD}}(f)$$

- Parameters: $(M, q, A, \Delta t, \alpha)$

- Whittle likelihood, single-detector

- Uniform priors

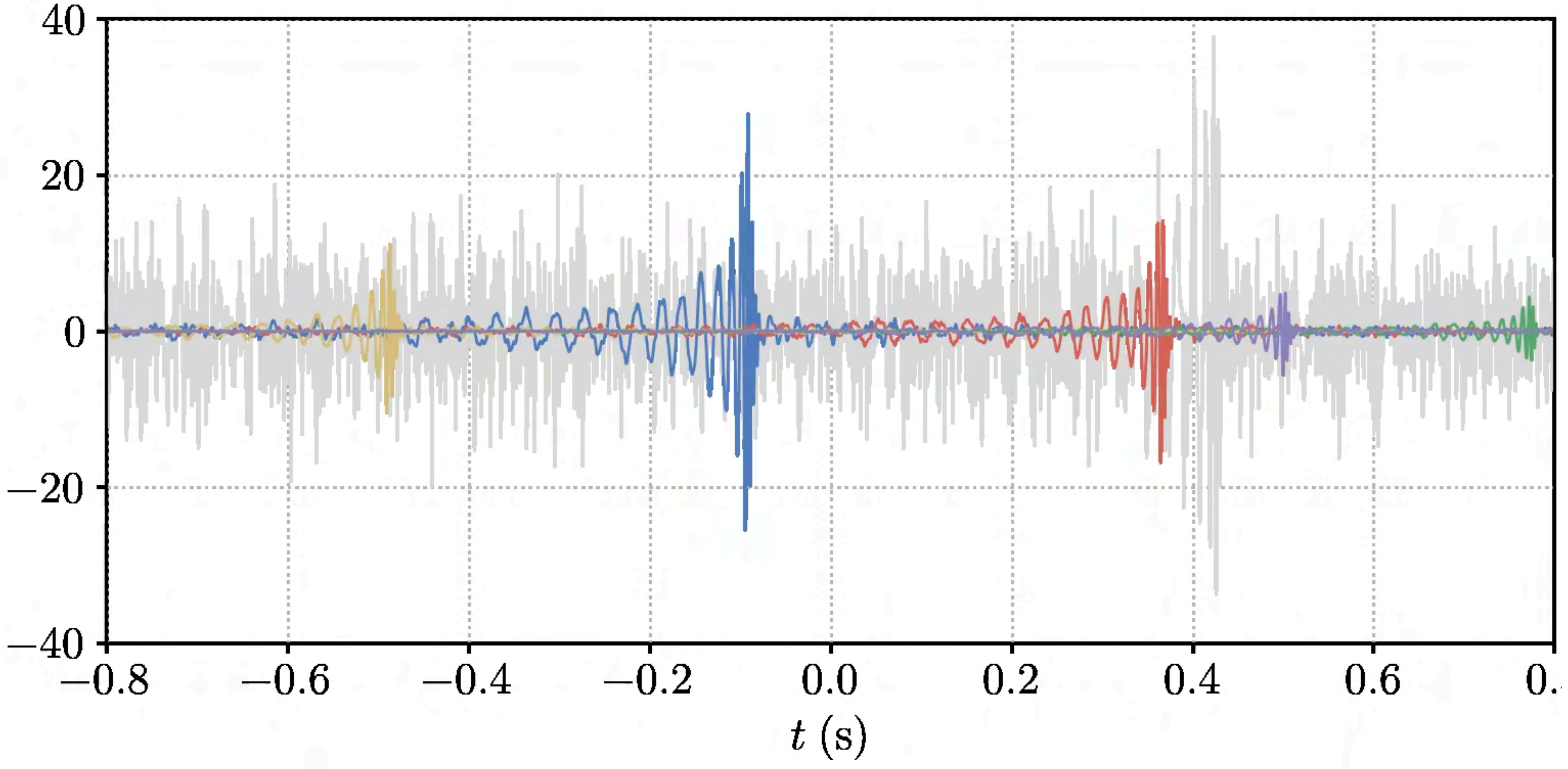
Note: simple amplitude and phase factor, replacing extrinsic parameters

$$(d_L, \iota, \varphi, \mathrm{ra}, \mathrm{dec}, \psi)$$

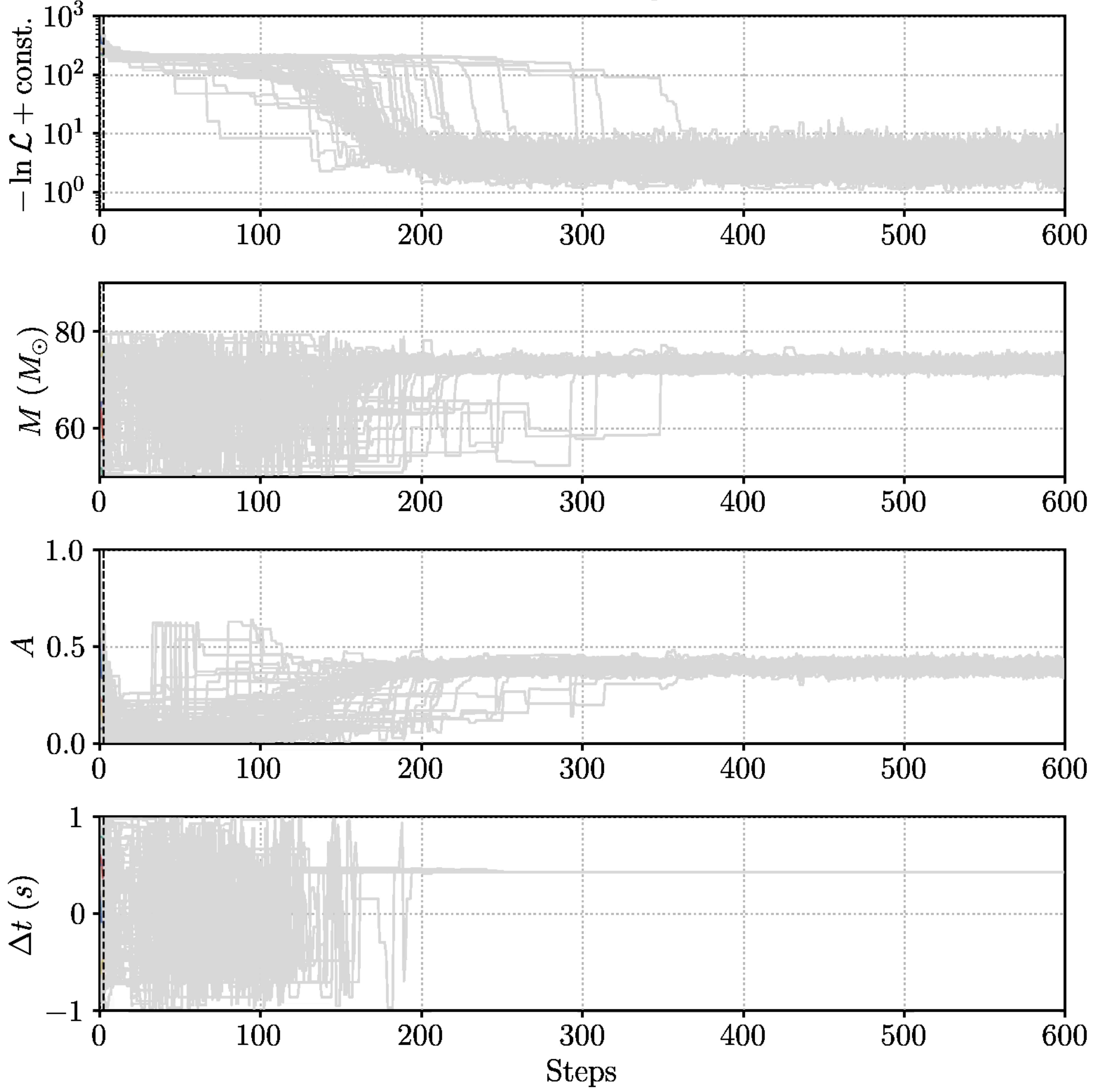
Walkers: 64
Temps: 5
Iters: 1000
Time: 15min

MH MCMC example: result

Sample waveforms (5 random walkers)

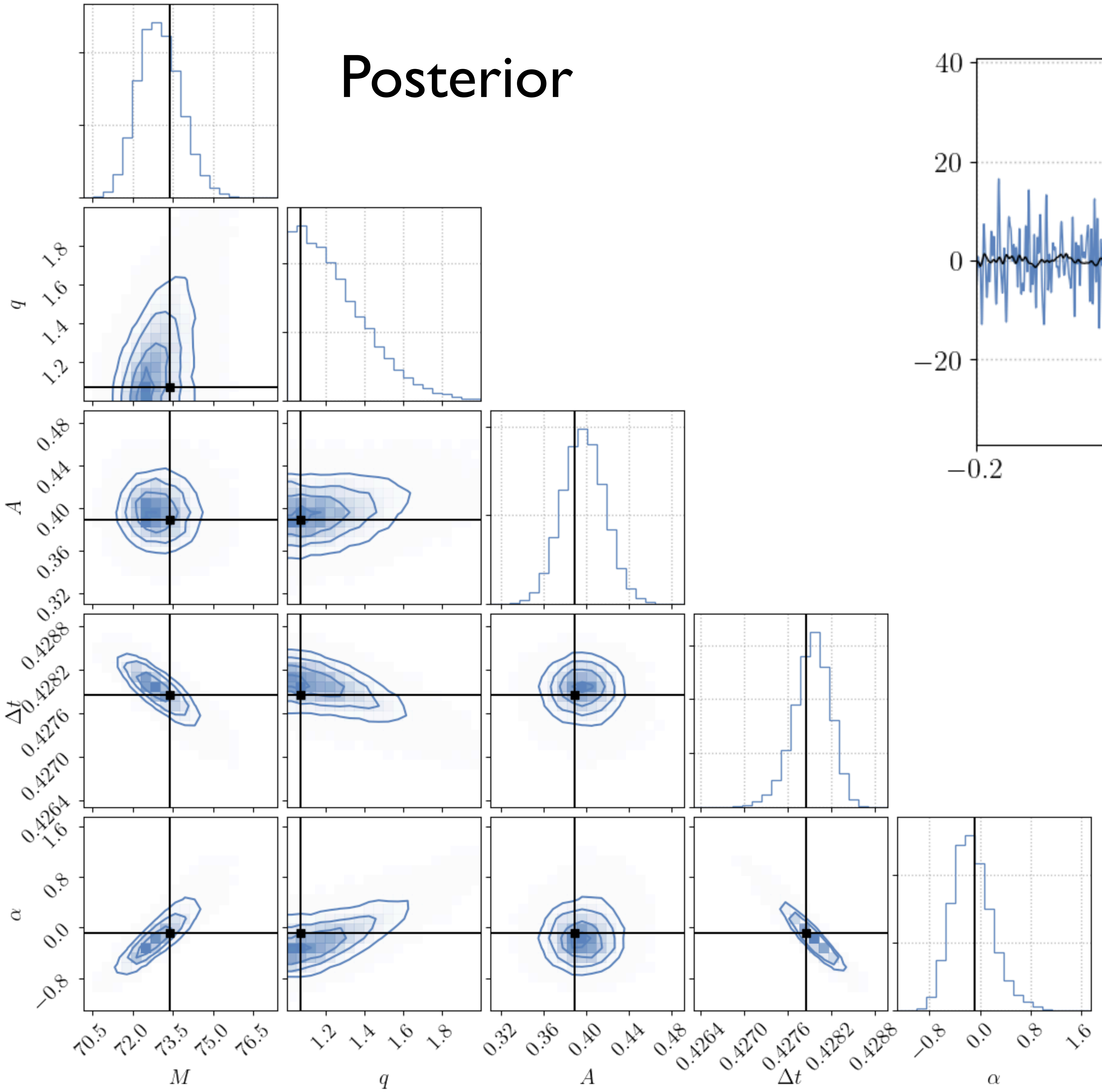


Trace plot

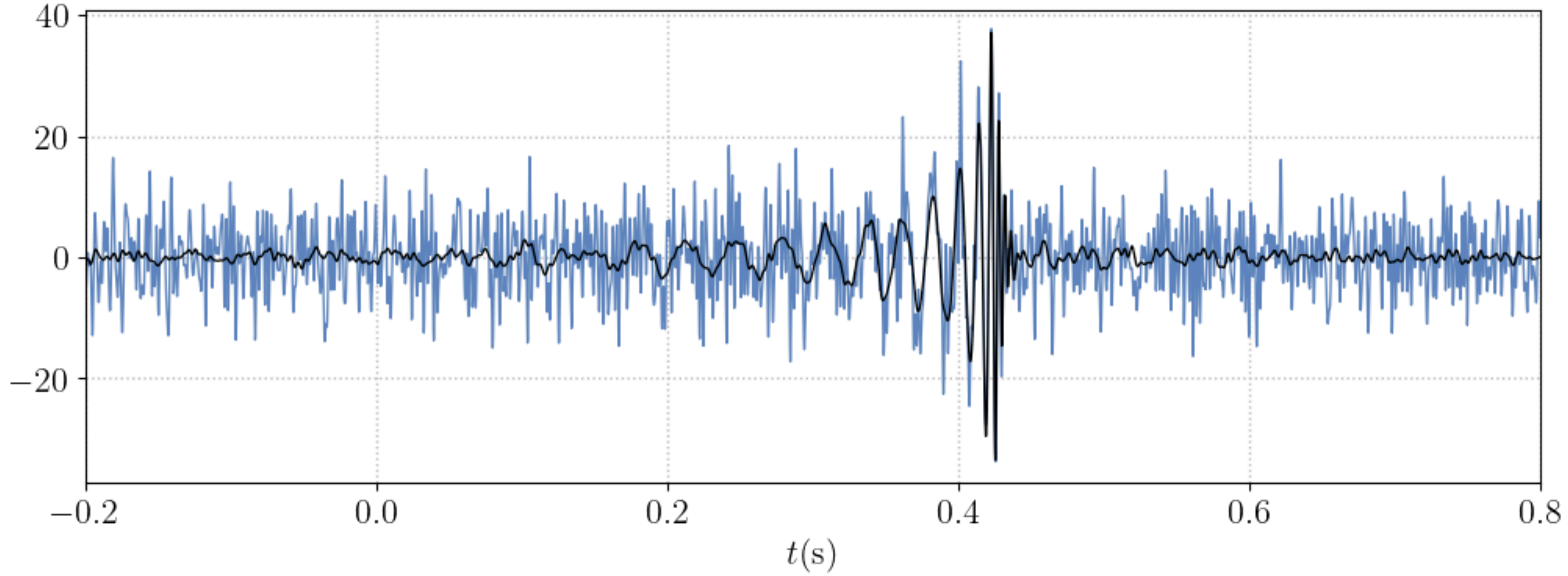


MH MCMC example: result

Posterior



Max-likelihood waveform



Outline

Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

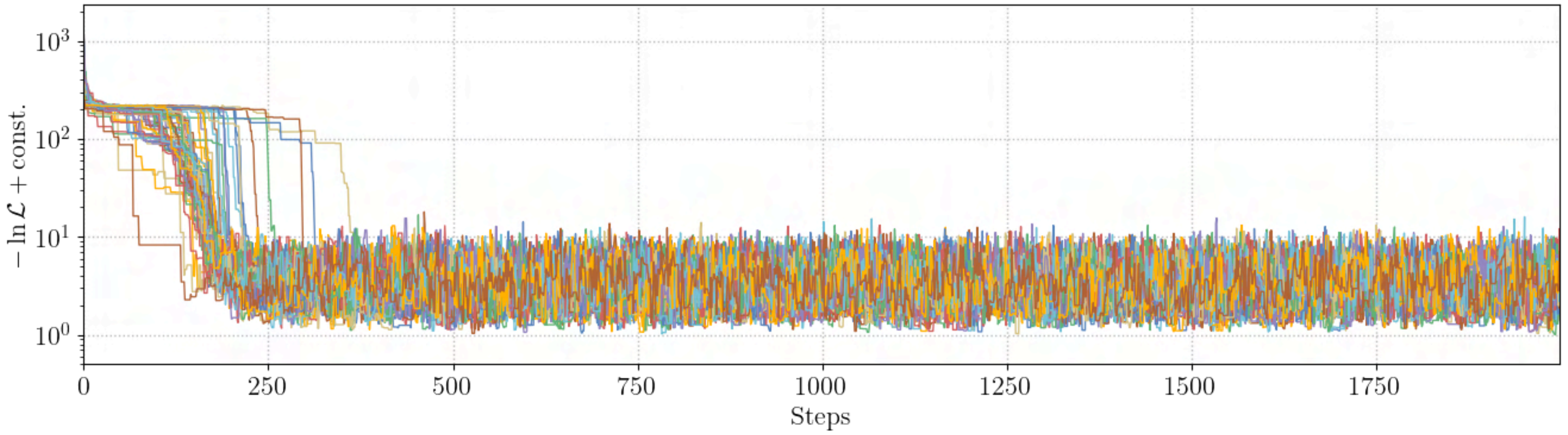
Part II

- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- **PE toolbox**
- PE results from LVK
- Future detectors and their challenges

Qualifying PE results: convergence

Trace plots

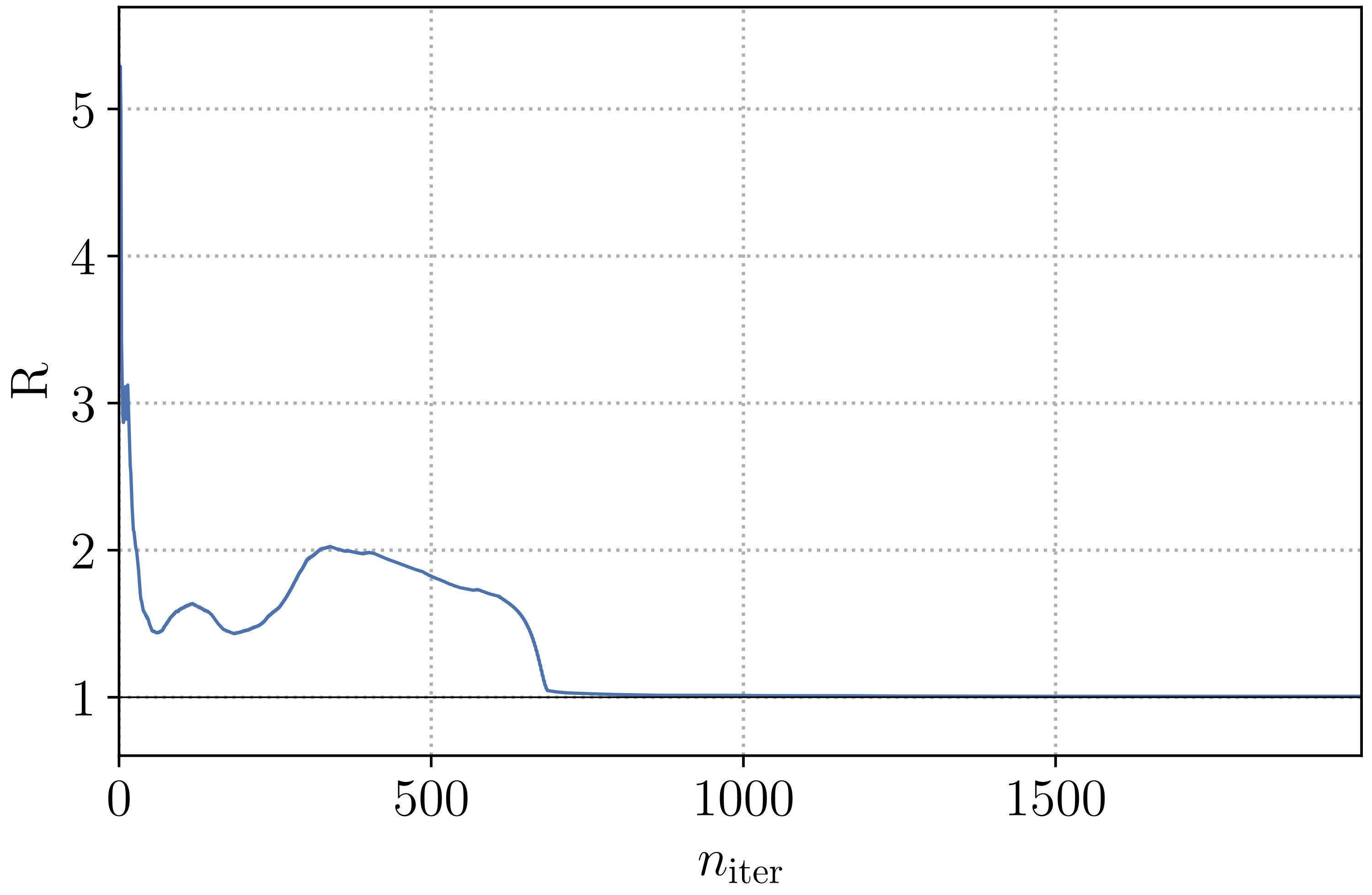
Help identify burn-in phase



Gelman-Rubin

R: in-chain and between-chain variance

Should have R=1



Autocorrelation length

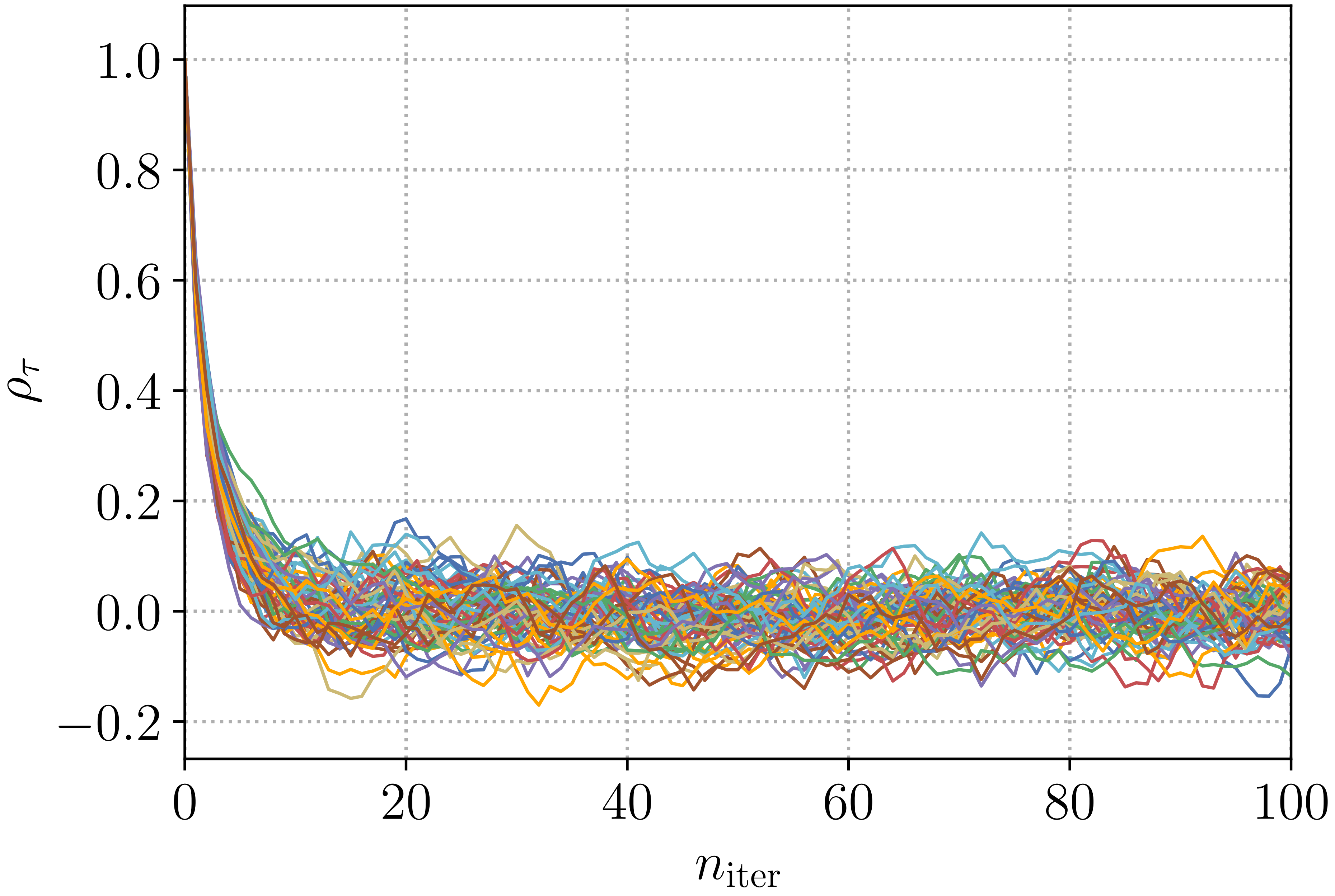
Autocorrelation:

$$\hat{c}_f(\tau) = \frac{1}{N-\tau}\sum_{n=1}^{N-\tau}(f_n-\mu_f)(f_{n+\tau}-\mu_f)$$

Autocorrelation length:

$$\hat{\rho}_f(\tau) = \hat{c}_f(\tau)/\hat{c}_f(0) \qquad \tau_f = \sum_{\tau=-\infty}^{\infty}\rho_f(\tau)$$

Effective samples: N/τ_f



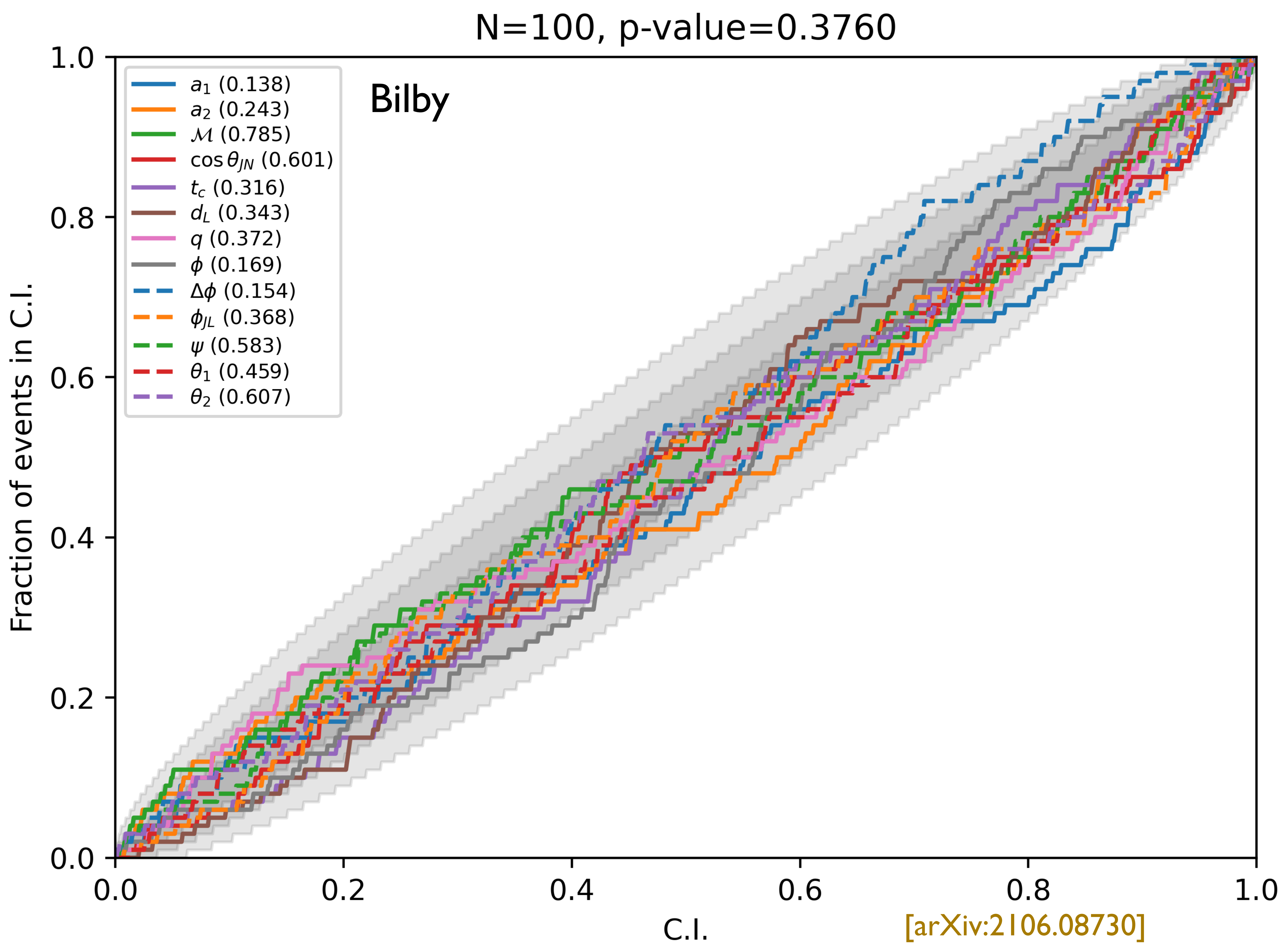
Qualifying PE results: quantile-quantile plots

- Idea: the true value must be in the x% confidence interval x% of the time

- Simulate a number of PE runs, different noise realizations and systems

- Expected deviations from unity known:

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

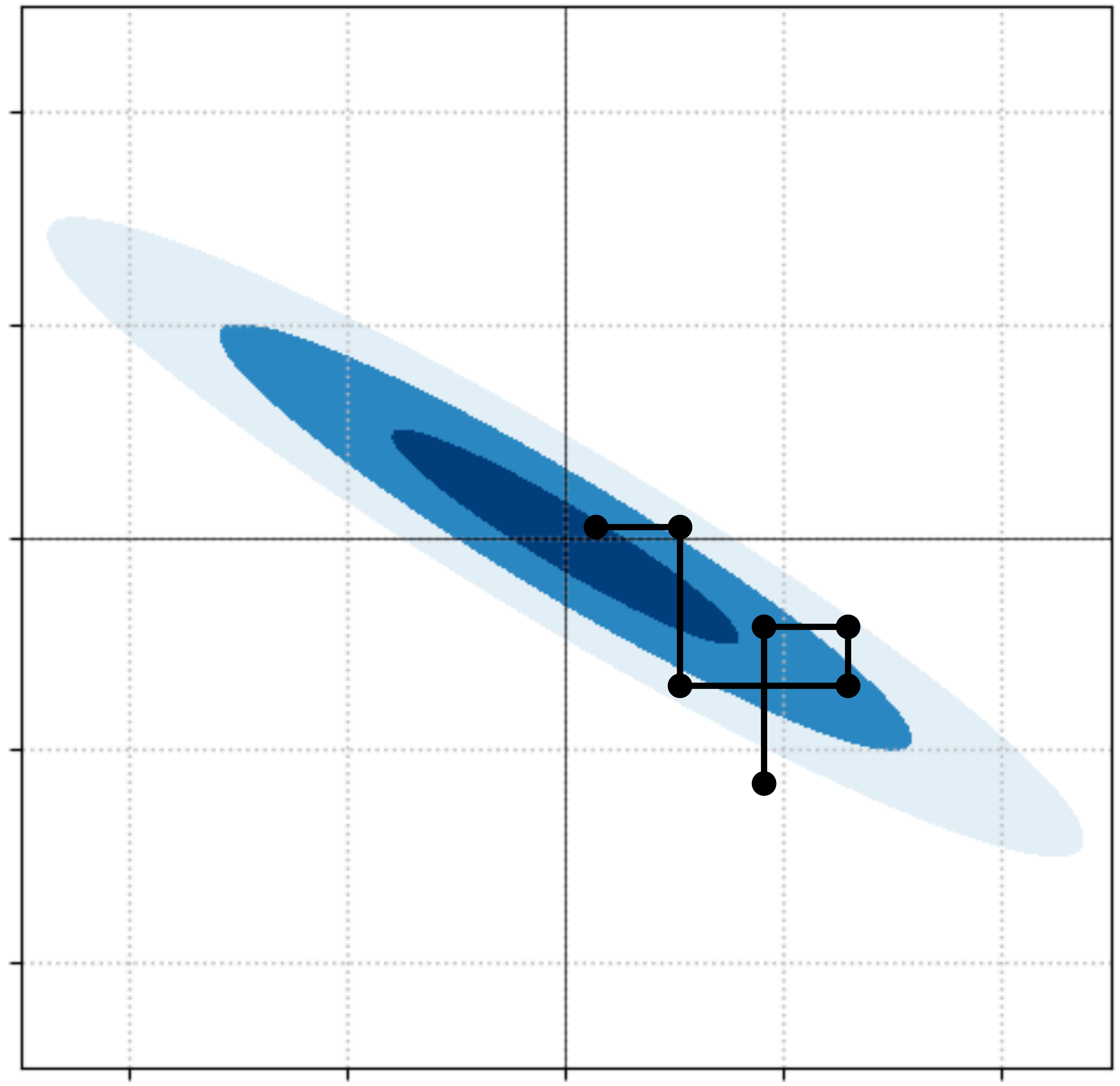


[arXiv:2106.08730]

Gibbs sampling

Update successively parameters:

$$x_{i+1} \sim p(x|y_i)$$
$$y_{i+1} \sim p(y|x_{i+1})$$



Usage

- decomposing between fast and slow parameters
- sampling across superposed sources
- caveat: inefficient with strong correlations

Elimination of extrinsic parameters and F-statistic

Likelihood marginalization

- Marginalize over time $$\int df/S_n\, e^{2i\pi f\Delta t}\tilde{h}\tilde{d}^* \rightarrow \mathrm{IFFT}[\tilde{h}\tilde{d}^*]$$
- Marginalize over phase (not possible with HM) $$\int d\phi_0 \, \exp\left[\int df/S_n\, e^{i\phi_0}\tilde{h}\tilde{d}^*\right] \rightarrow I_0\left[\left|\int df/S_n\, \tilde{h}\tilde{d}^*\right|\right]$$

Likelihood, not log-likelihood !

[arXiv:1409.7215]

Likelihood optimization (F-stat)

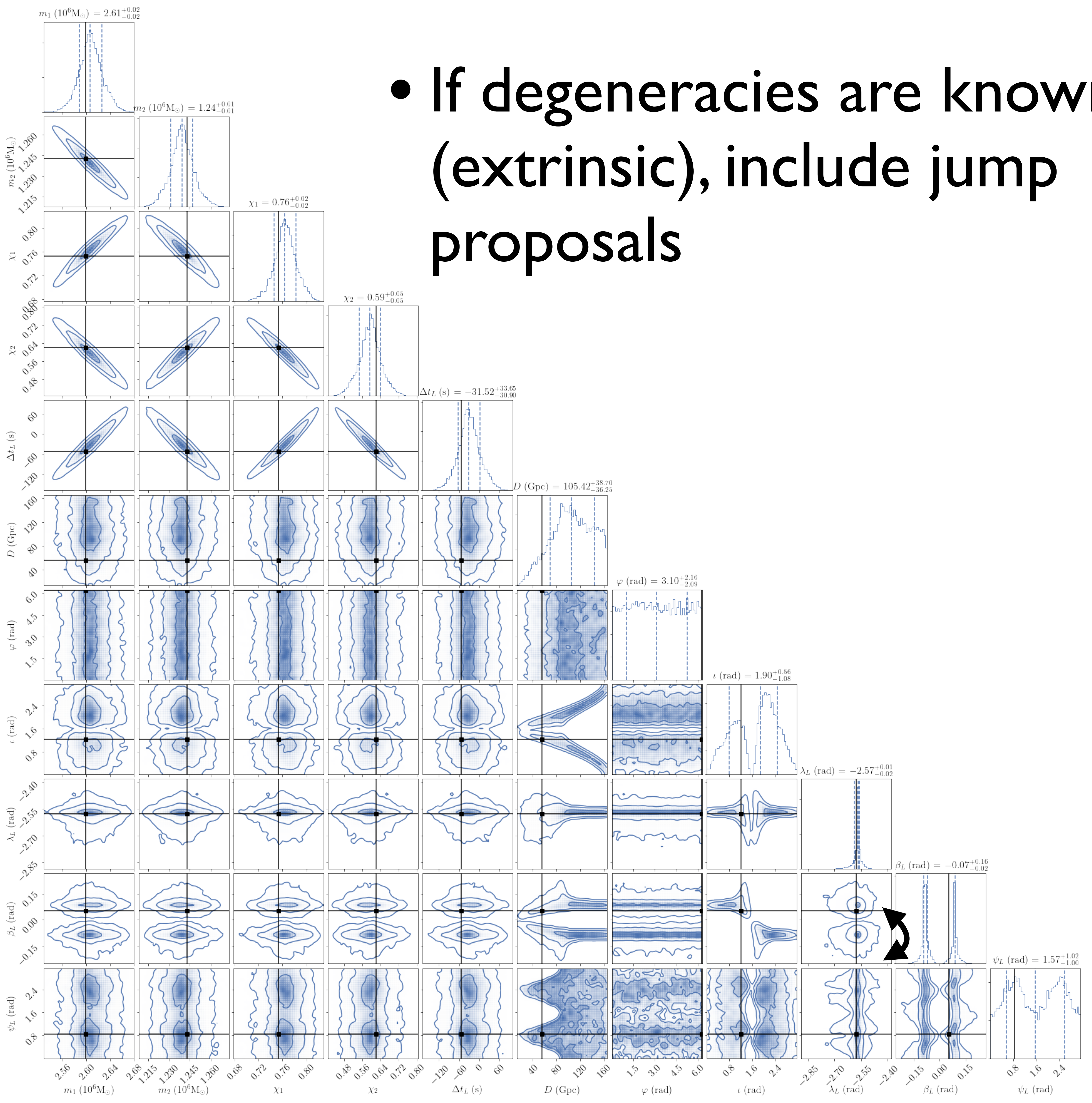
- If quantities affect linearly the signal, loglikelihood is quadratic in them and optimization is simple

$$\ln\mathcal{L} = -\frac{1}{2}(Ae^{i\alpha}h - d|Ae^{i\alpha}h - d)$$

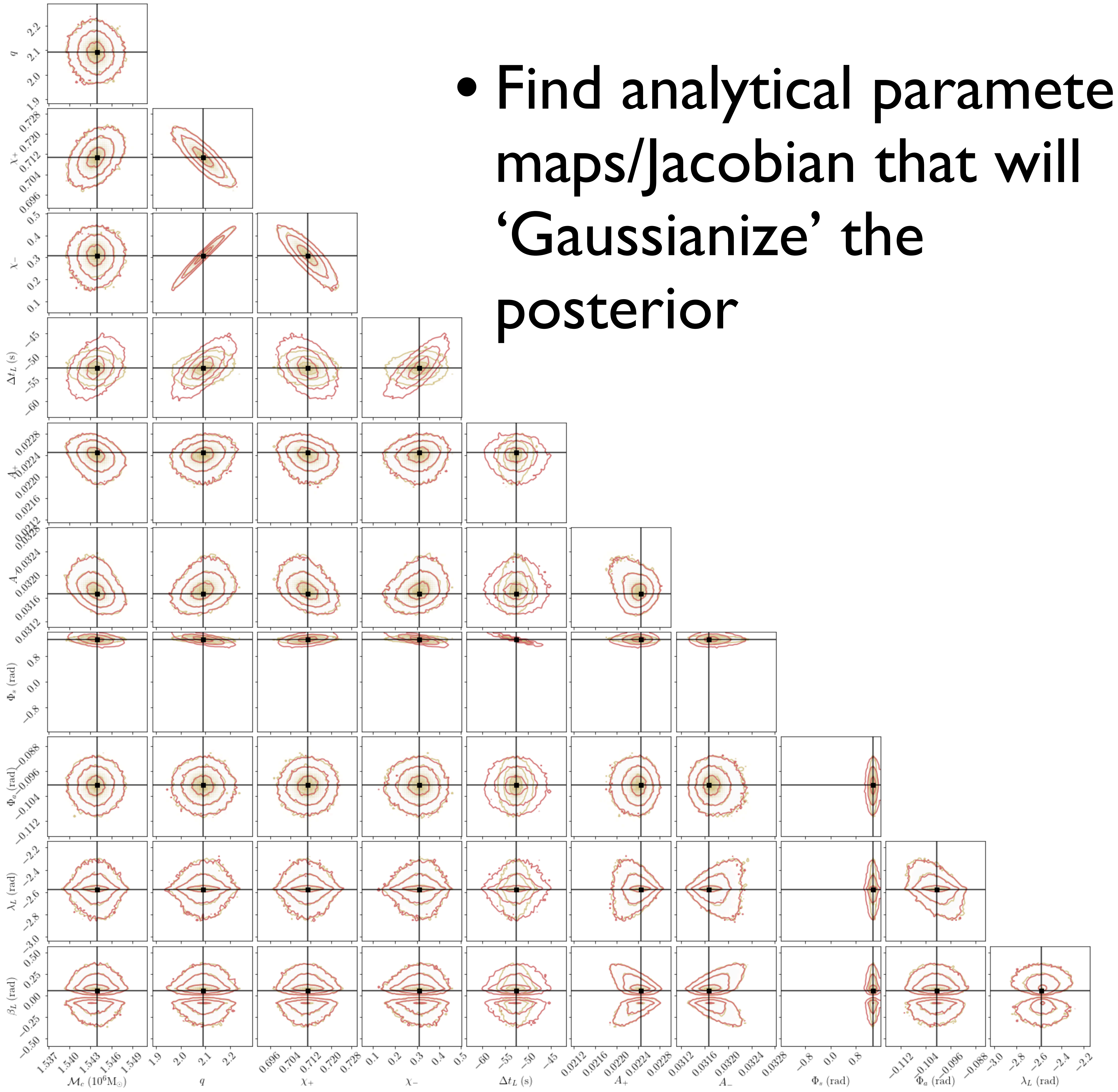
$$\frac{\partial \ln\mathcal{L}}{\partial A} = 0 \qquad \frac{\partial \ln\mathcal{L}}{\partial \alpha} = 0$$

- Not related to posterior, but very useful for search (reduced dimensions)

On the choice of parameters for sampling



- If degeneracies are known (extrinsic), include jump proposals



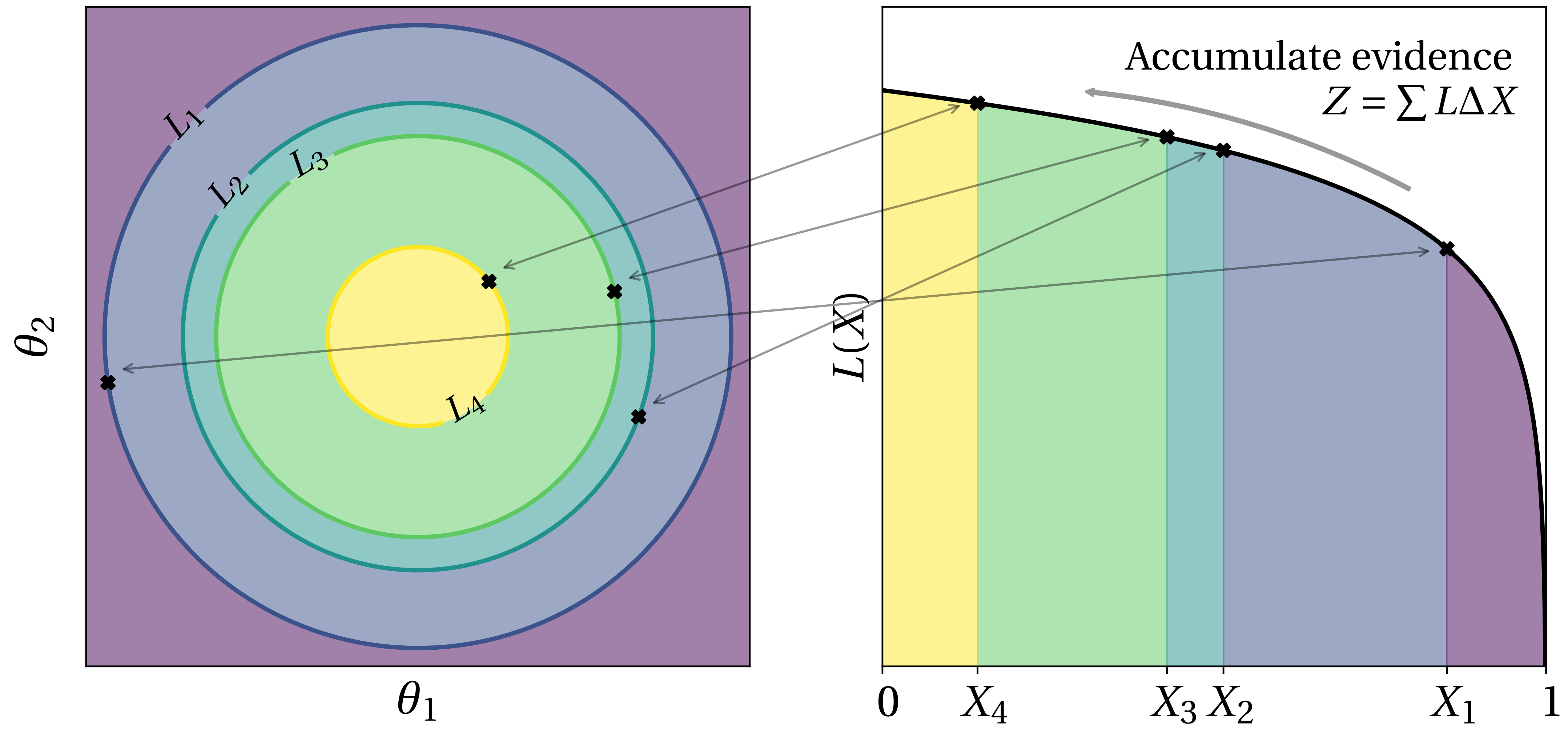
- Find analytical parameter maps/Jacobian that will 'Gaussianize' the posterior

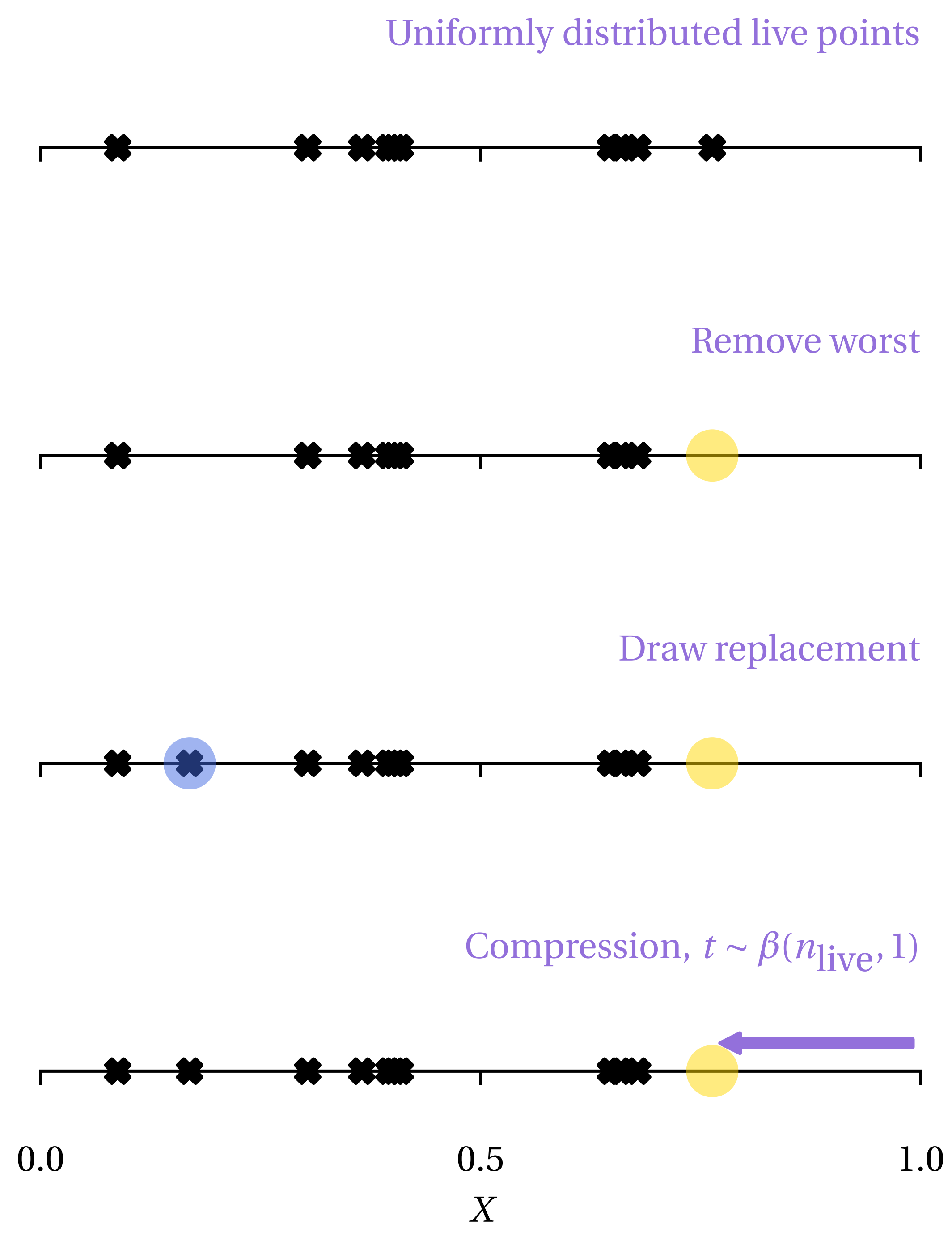
If possible, sample in what the detector observes !

Nested sampling

See Review [arXiv:2205.15570]

Compute evidence $p(d|M) = \int d\theta\, p(d|\theta, M) p(\theta|M)$
and obtain samples $\sim p(\theta|d, M)$





- Decompose space in isolikelihood contours, replace integral by 1D integral

$$X(L^\star) = \int_{L>L^\star} \pi(\Theta) \mathrm{d}\Theta \qquad Z = \int_0^1 L(X)\, \mathrm{d}X$$

- Introduce set of live points that will be iteratively replaced, weighted replaced points become posterior samples
- Sampling constrained prior: region sampling, step sampling

Panorama of codes

MCMC codes

- emcee [arXiv:1202.3665]
- ptemcee [arXiv:1501.05823]
- eryn [arXiv:2303.02164]

Codes for GW

- LALinference (MCMC, Nest) [arXiv:1409.7215]
- bilby (MCMC, dynesty) [arXiv:1811.02042]
- pyCBC inference [arXiv:1807.10312]

Nested sampling codes

- multinest [arXiv:0809.3437]
- polychord [arXiv:1506.00171]
- CPnest [https://github.com/johnveitch/cpnest]
- dynesty [arXiv:1904.02180]
- NessAI [arXiv:2102.11056]

Outline

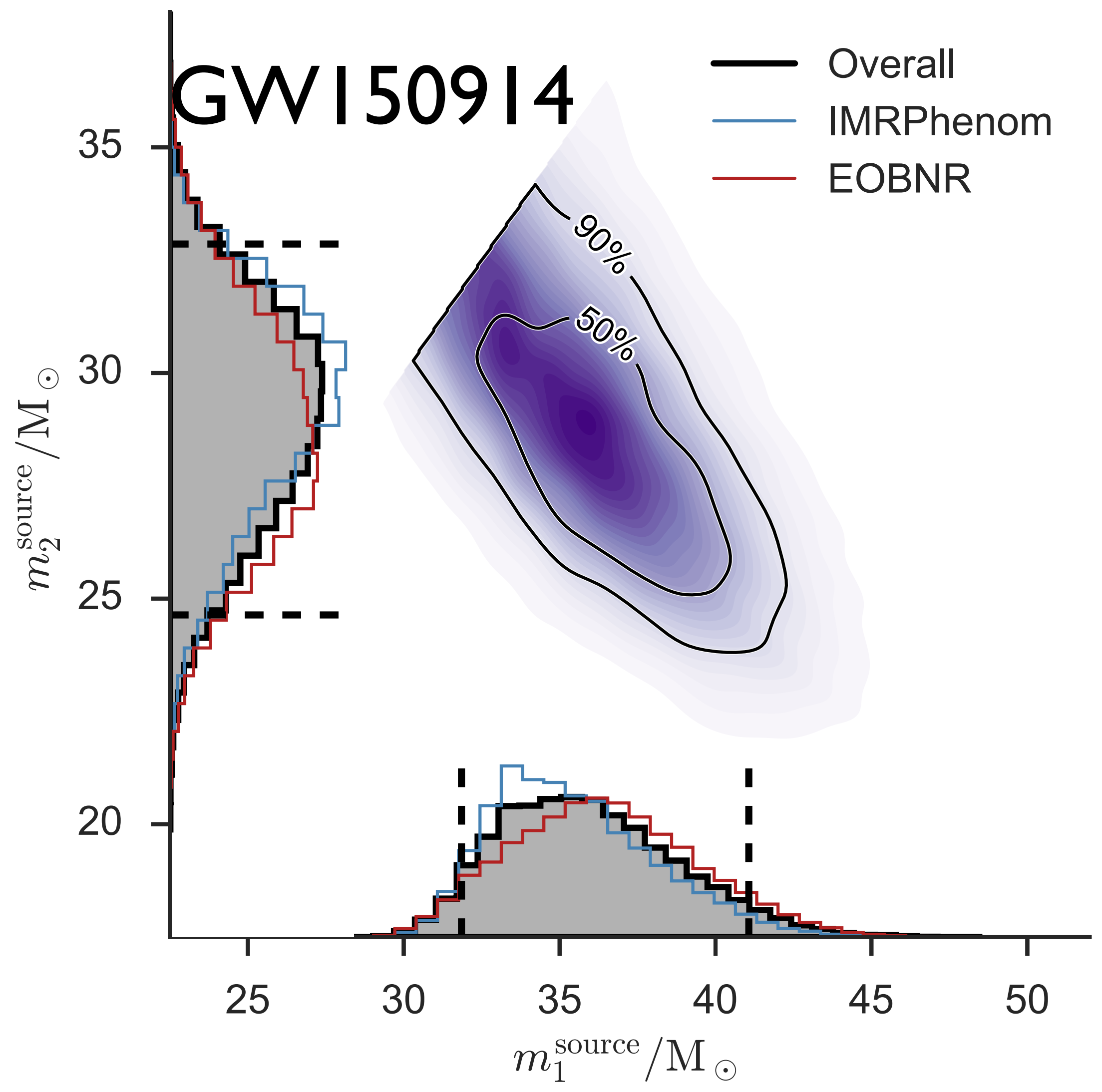
Part I

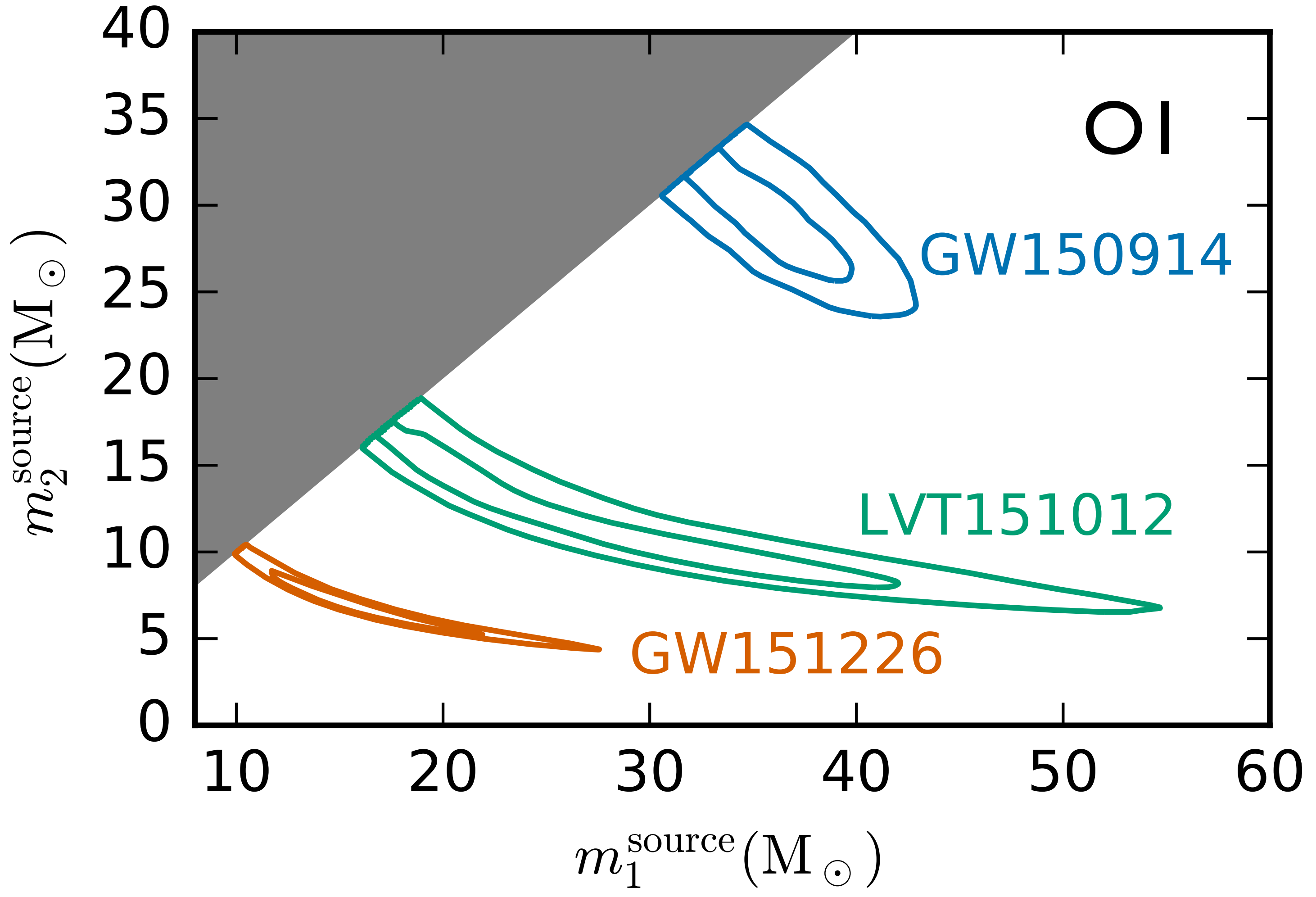
- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

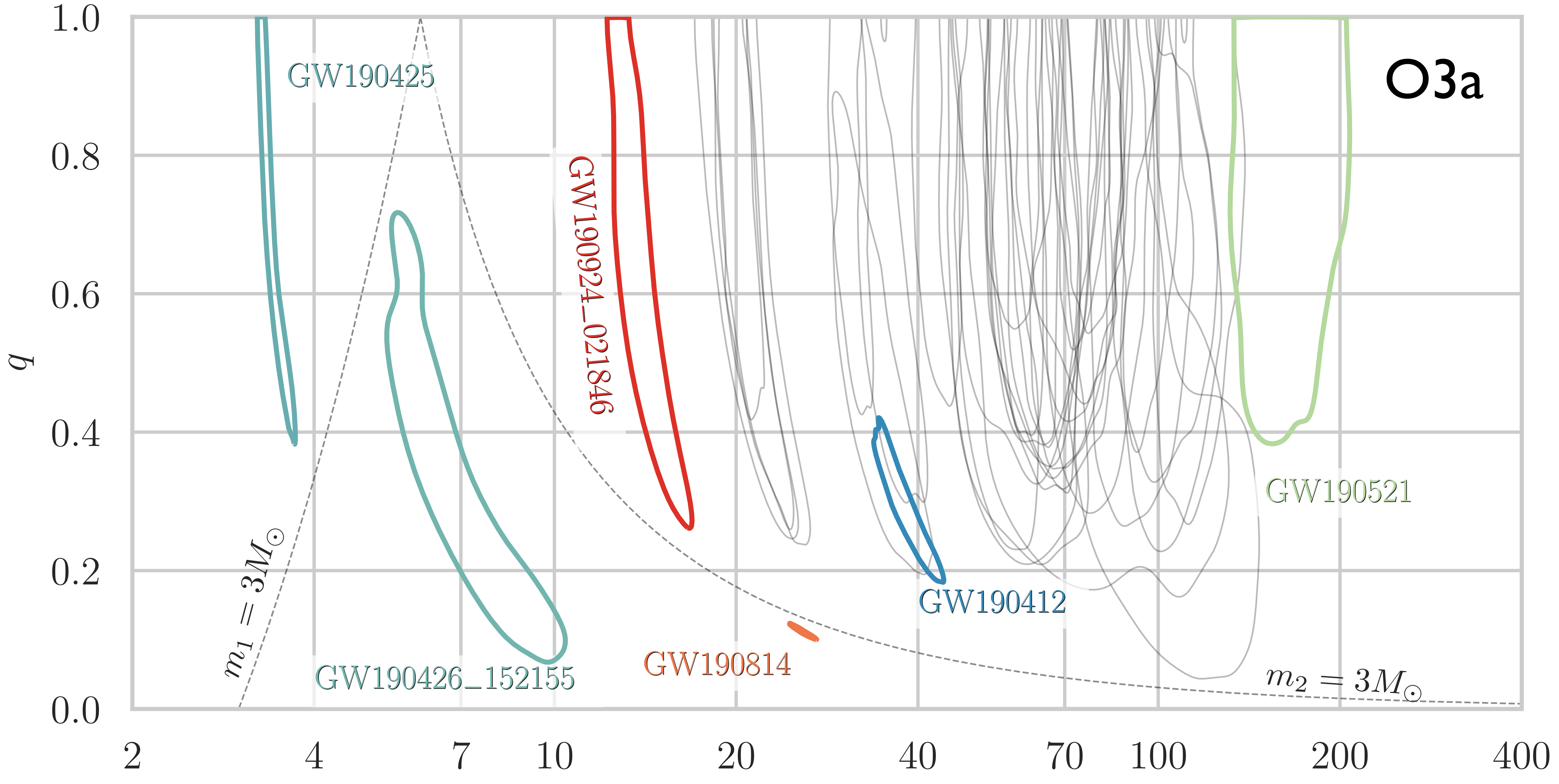
Part II

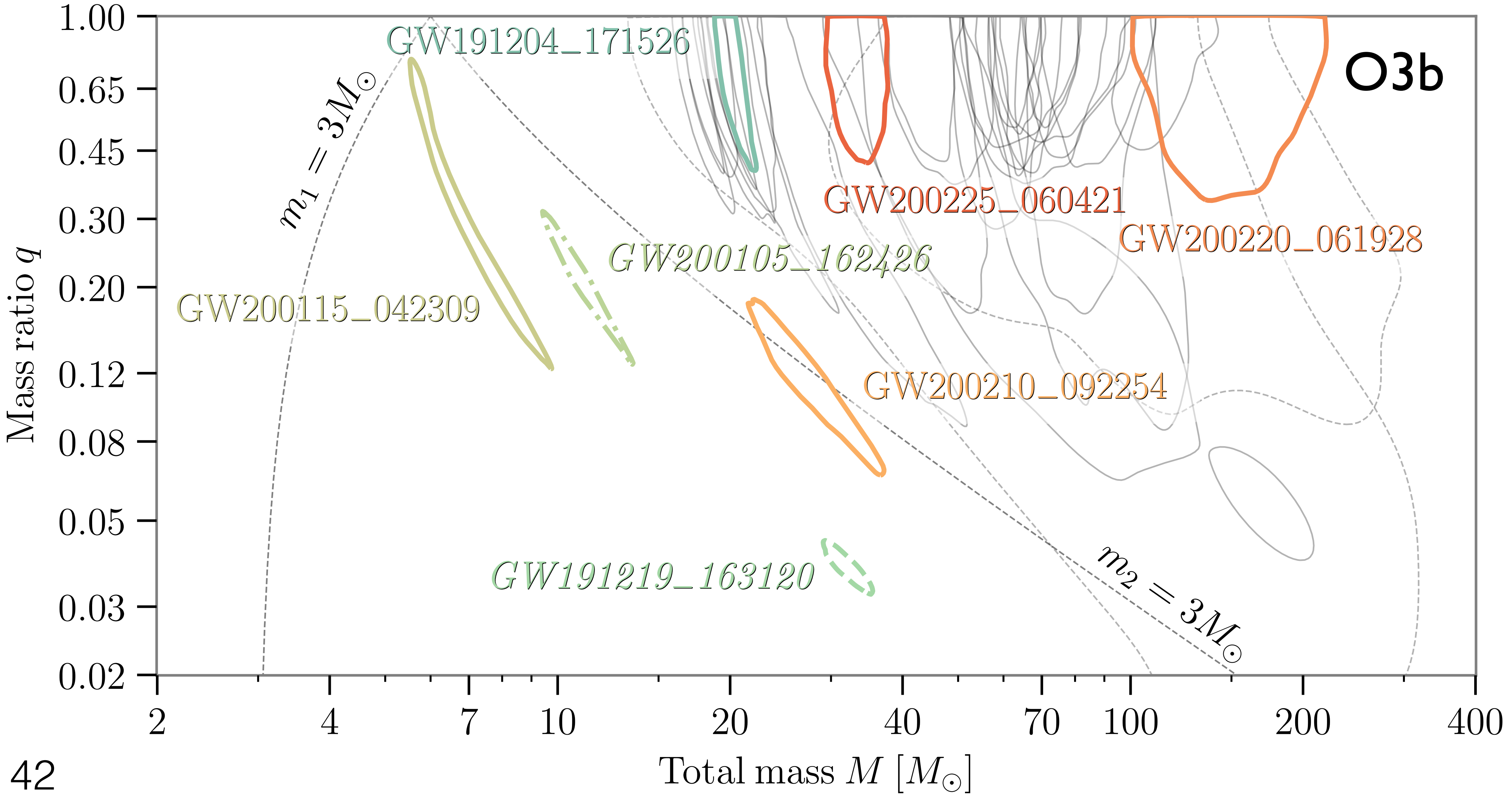
- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- **PE results from LVK**
- Future detectors and their challenges

PE results from LVK: mass posteriors



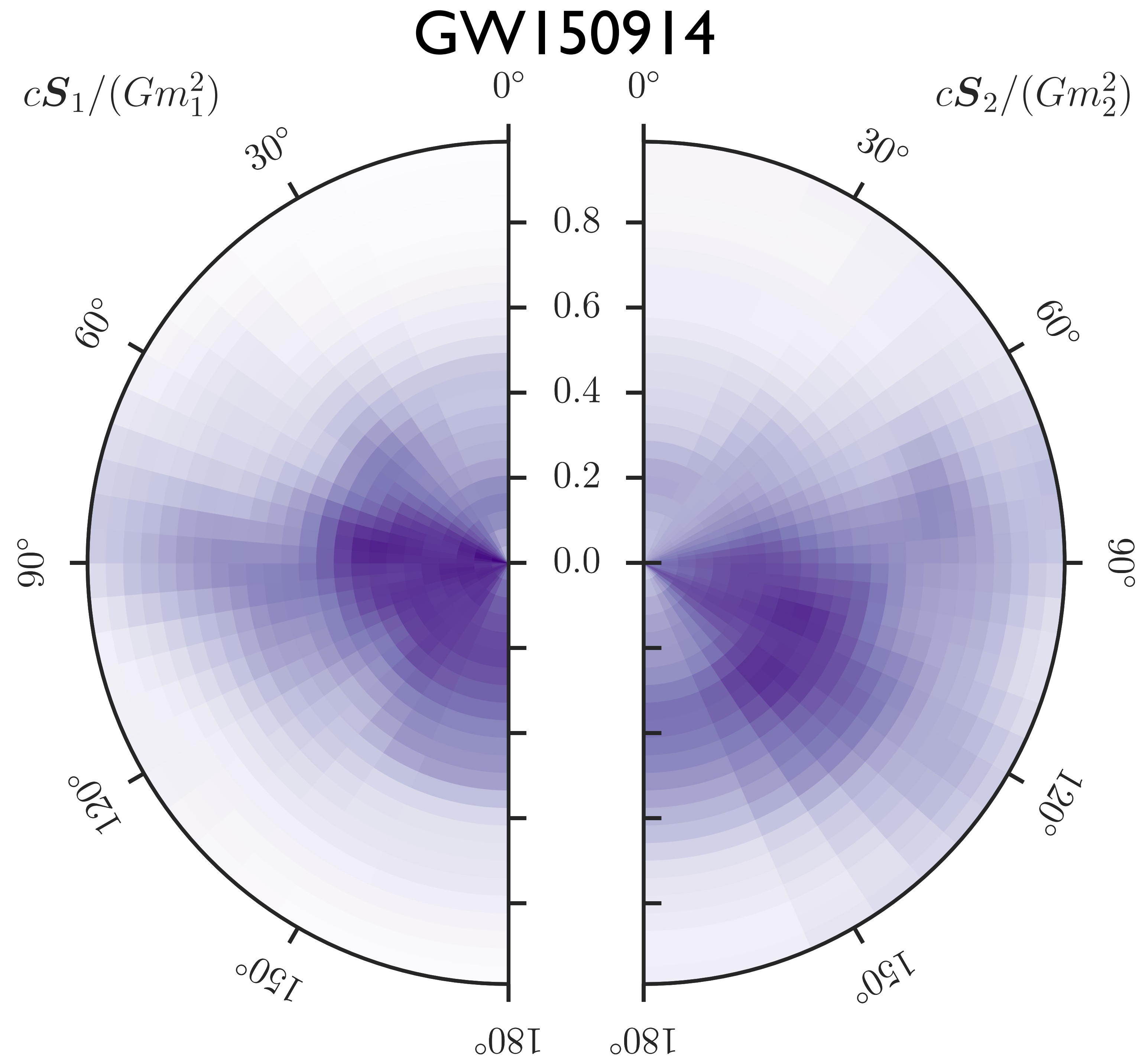


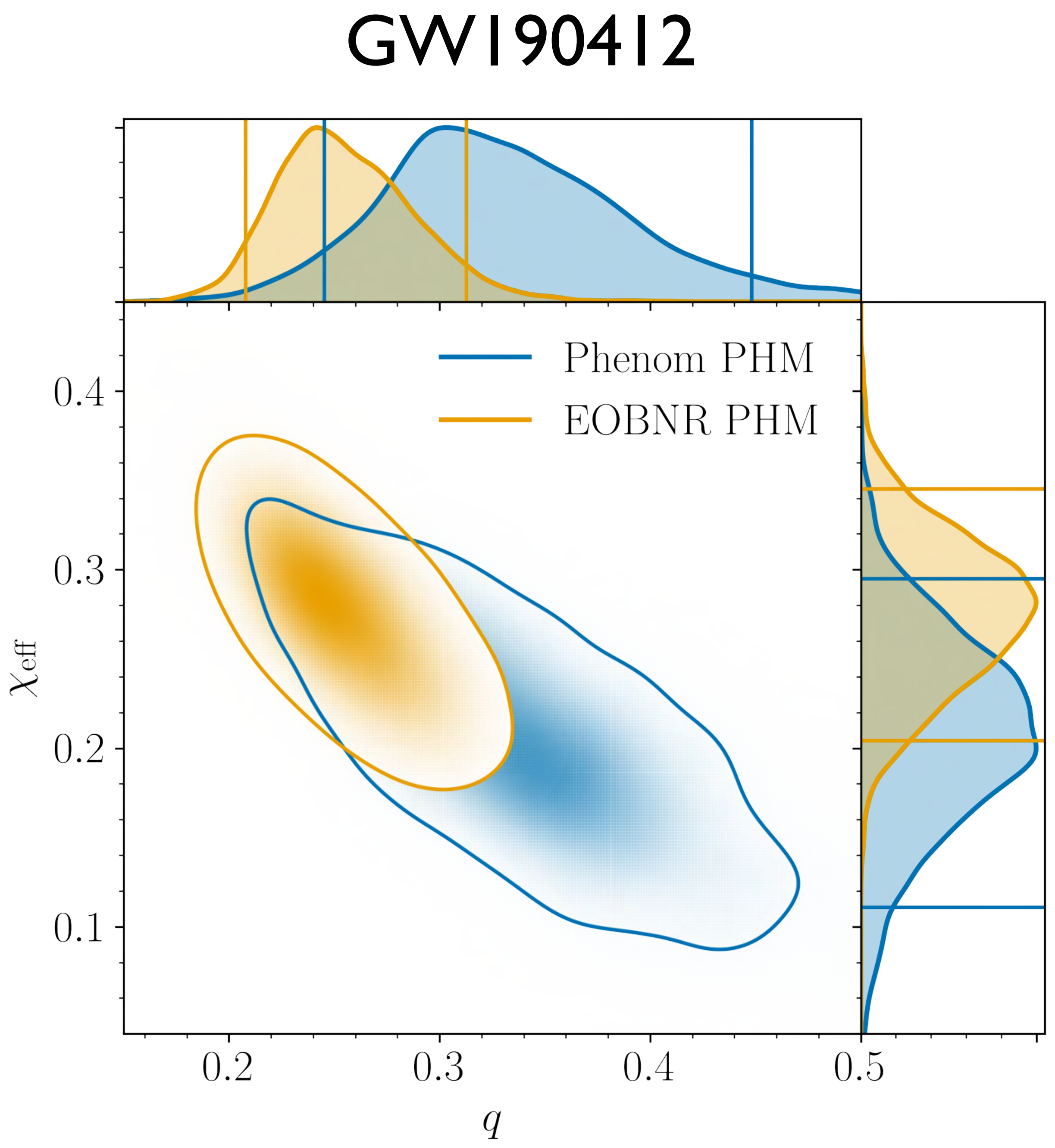


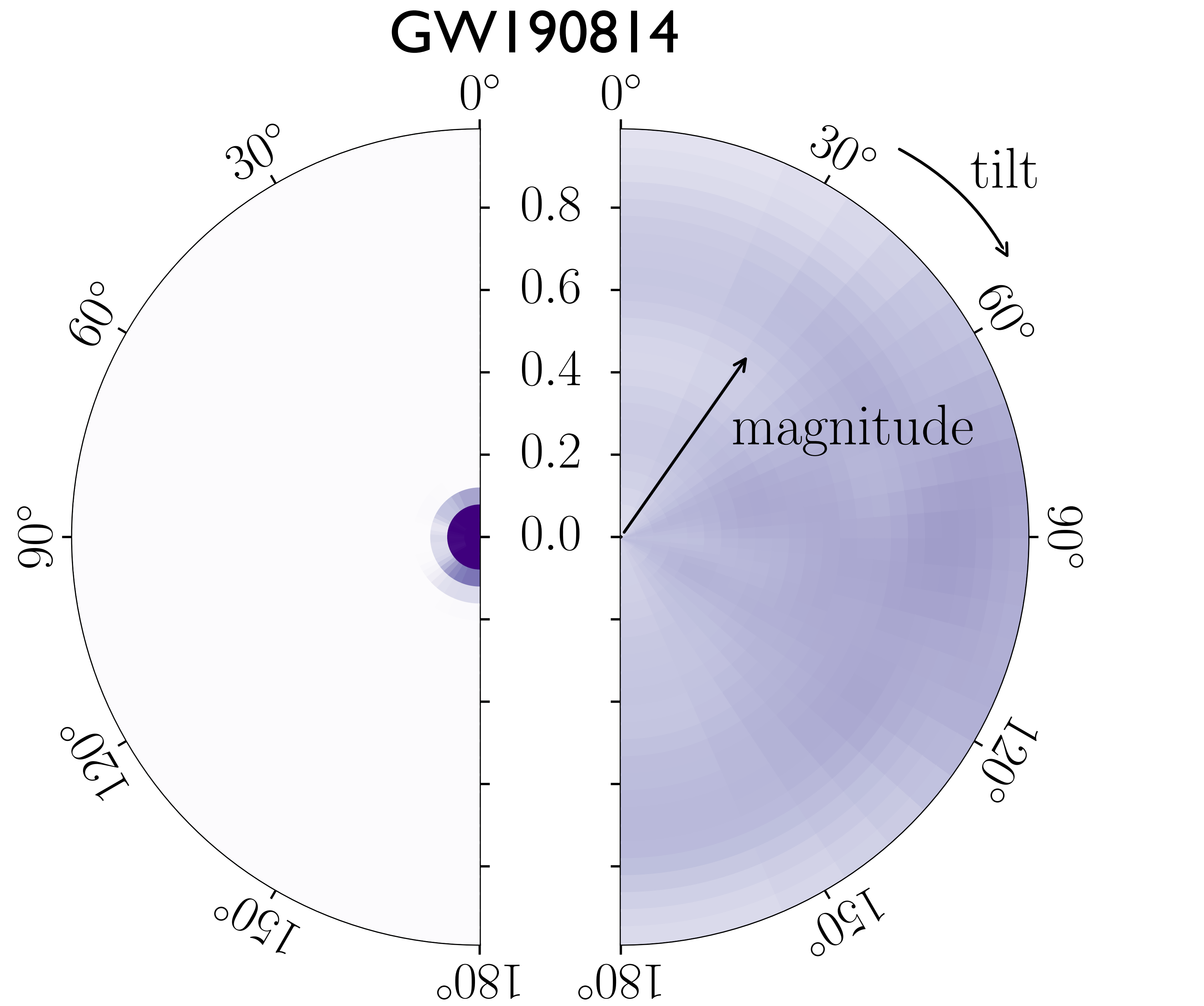


Chirp mass is best determined at low masses

PE results from LVK: spin posteriors

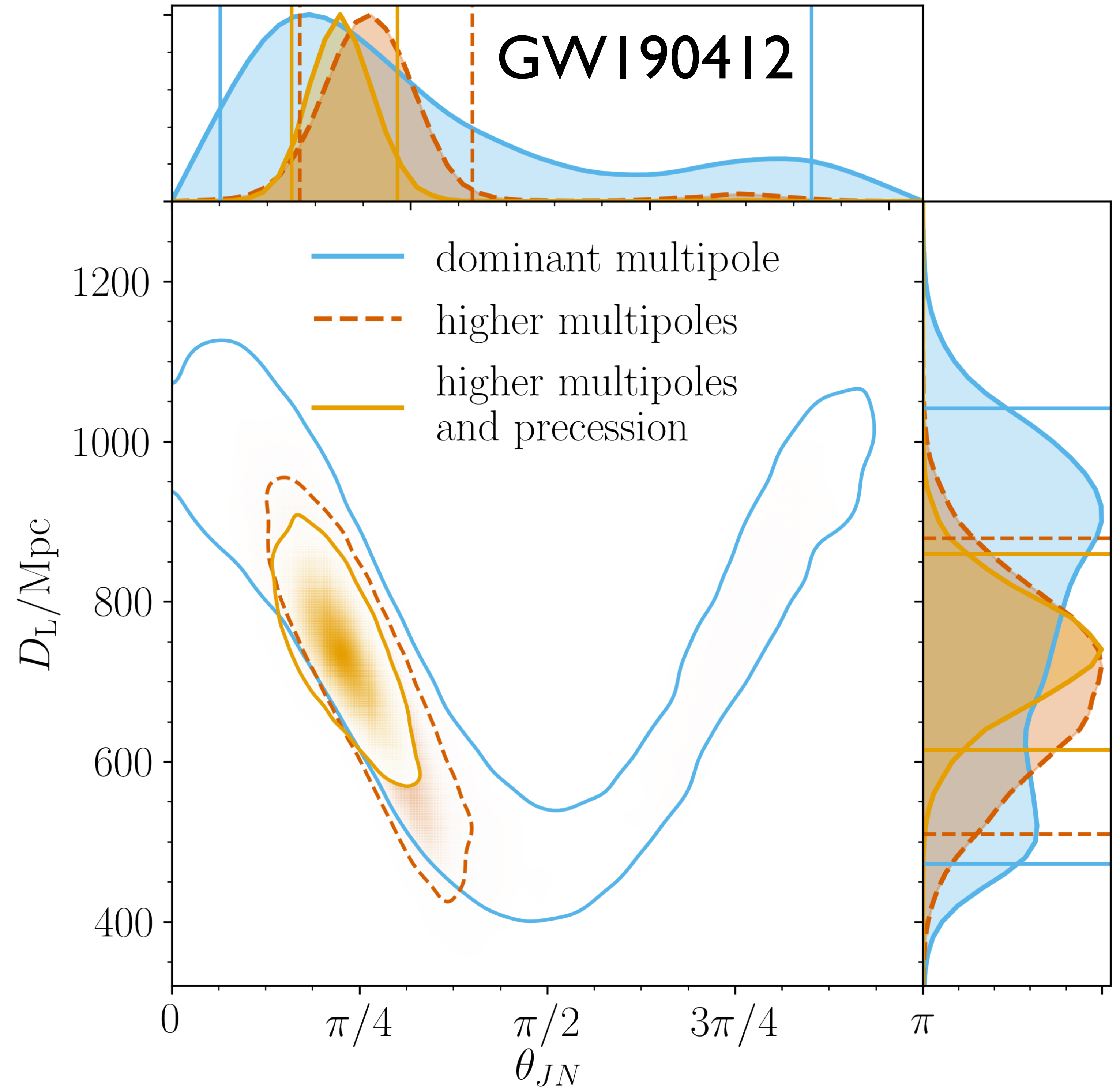


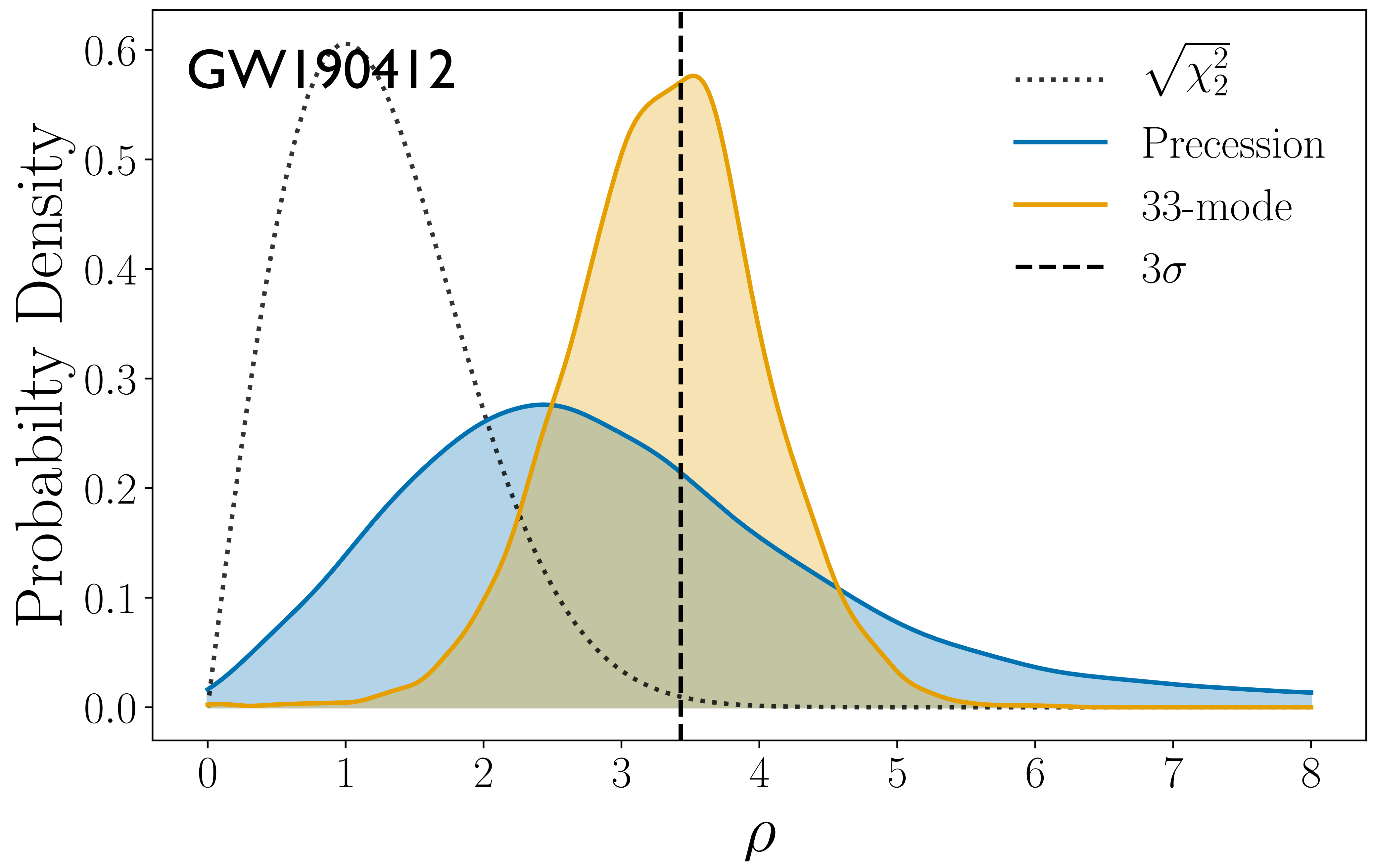




- Largely undetermined spins for many BBH events
- Aligned spin/mass ratio correlation
- A few detection of aligned spins
- NSBH events: strong constraints on primary spin

PE results from LVK: distance/inclination



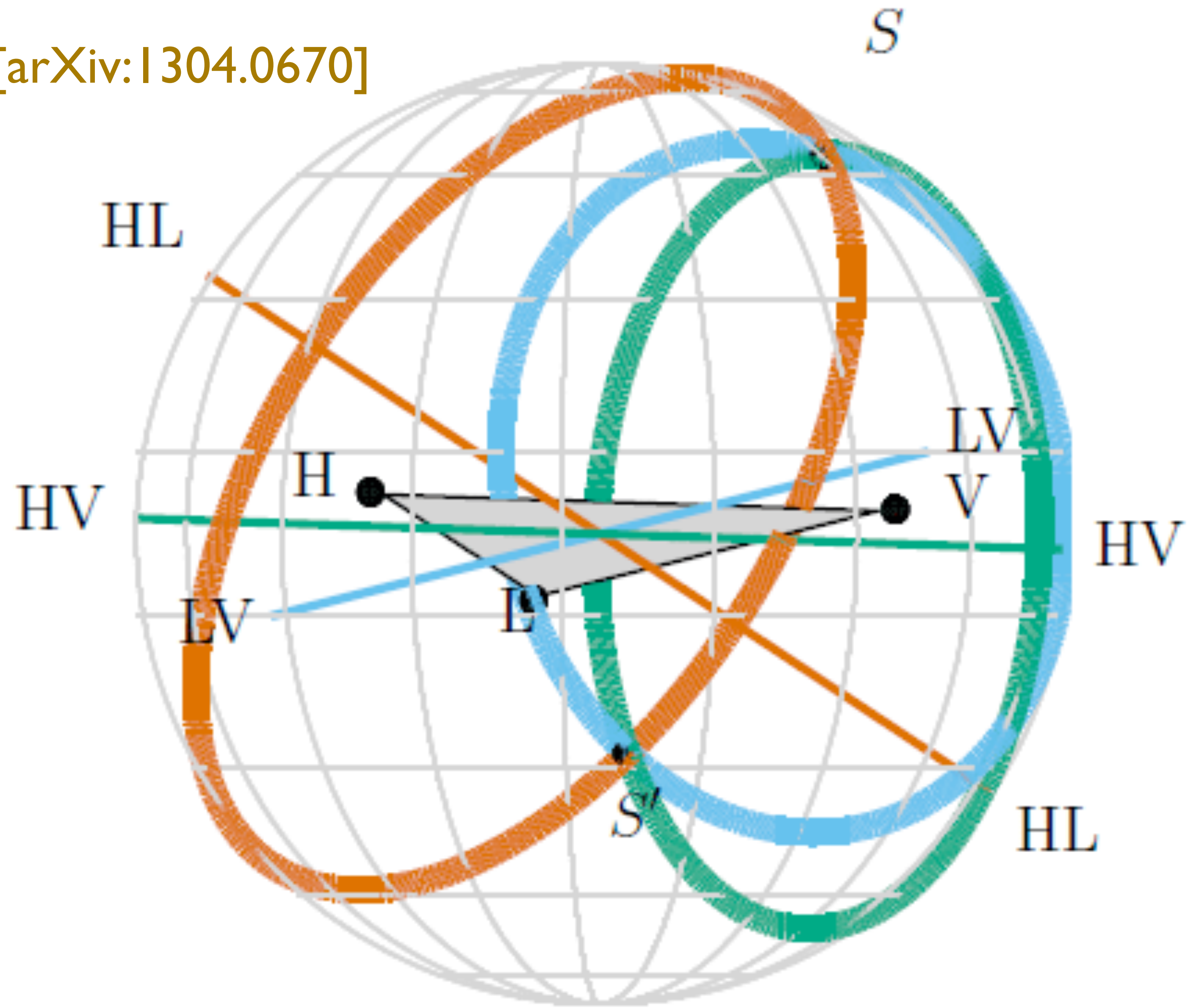


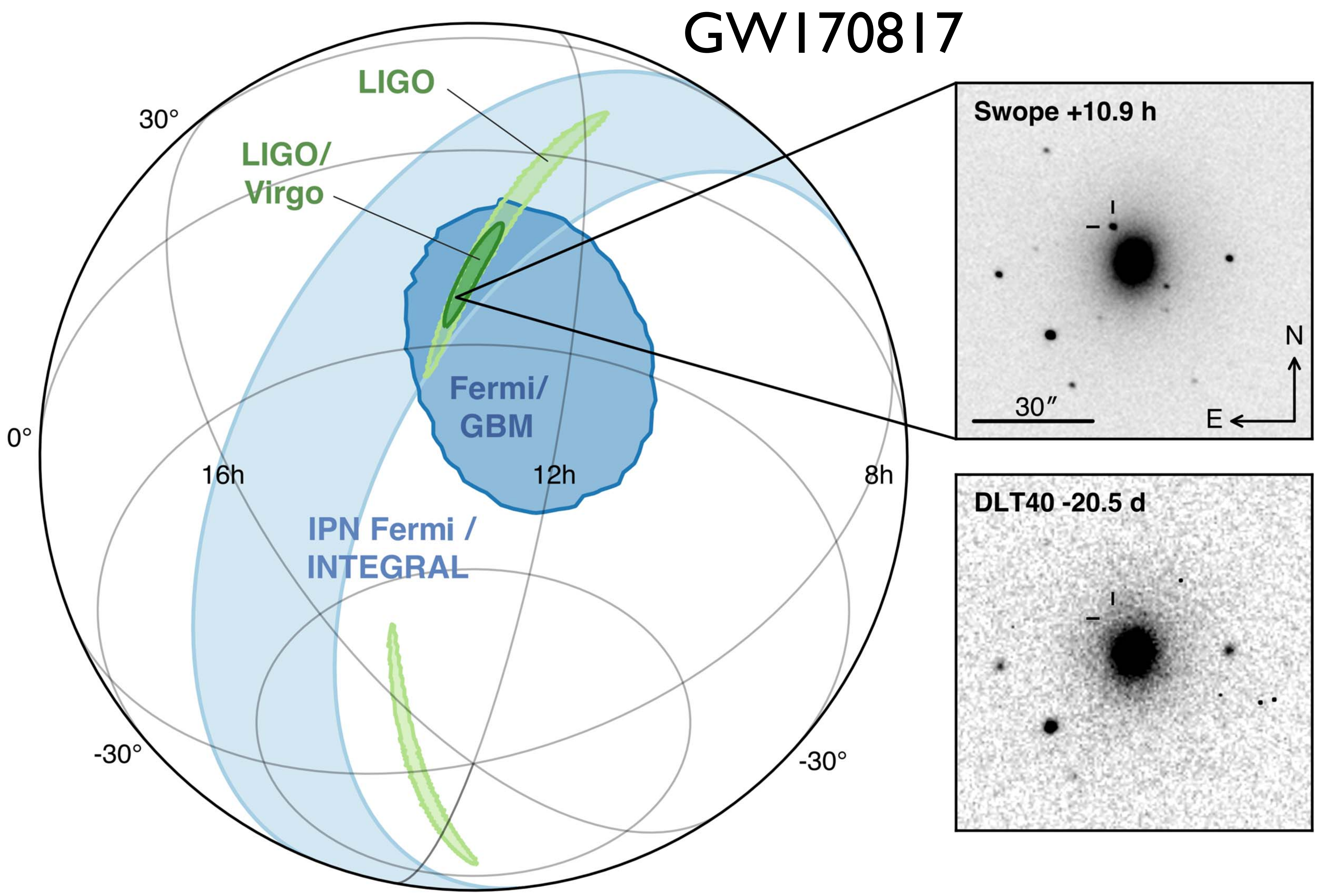
- GW190412: strong signal with high mass ratio q~4
- Distance-inclination degeneracy is broken by higher modes and precession

- Evidence for 33 mode in the data

Sky localization and rapid localization

[arXiv:1304.0670]





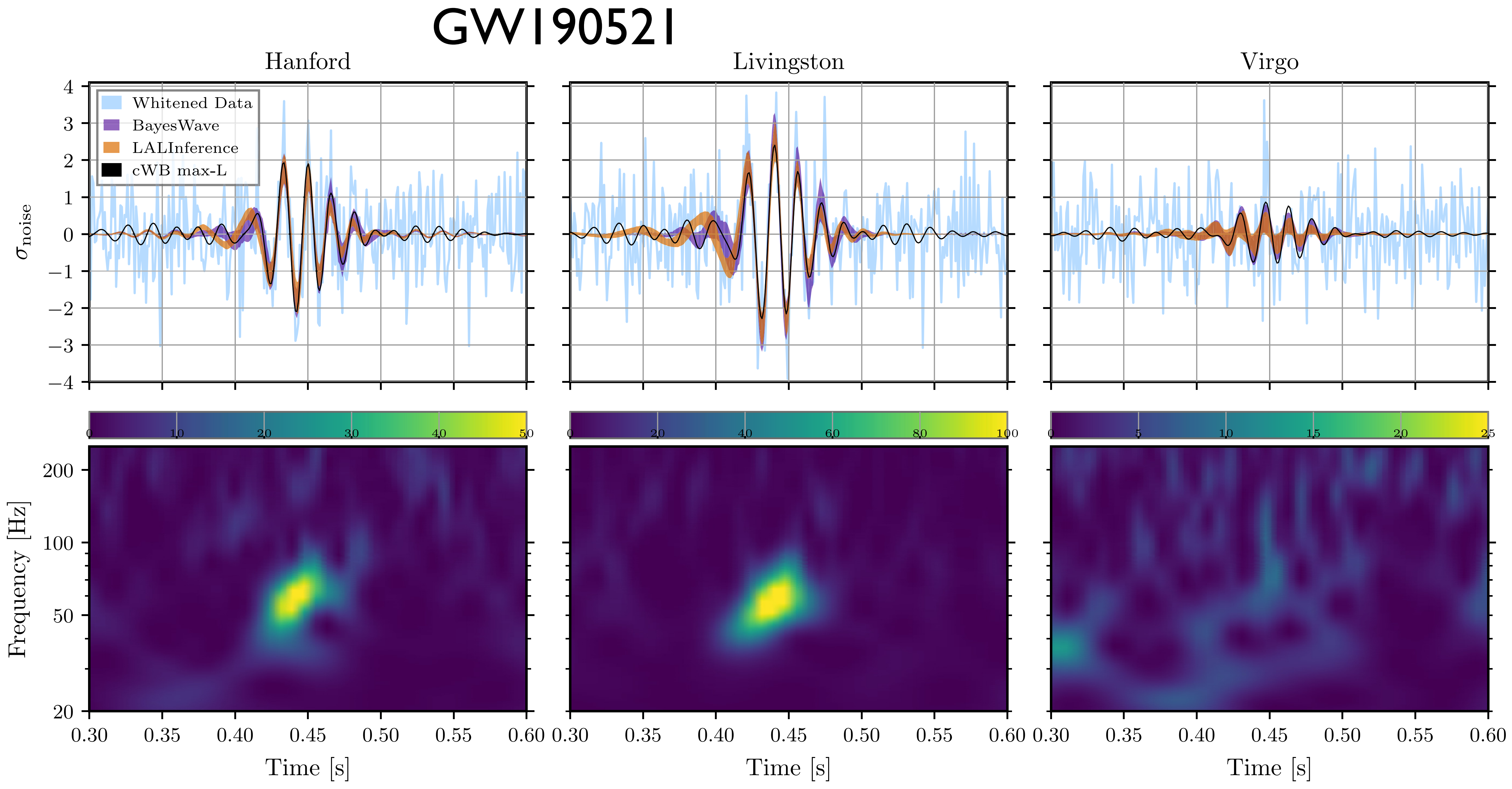
- Main information: triangulation by measuring

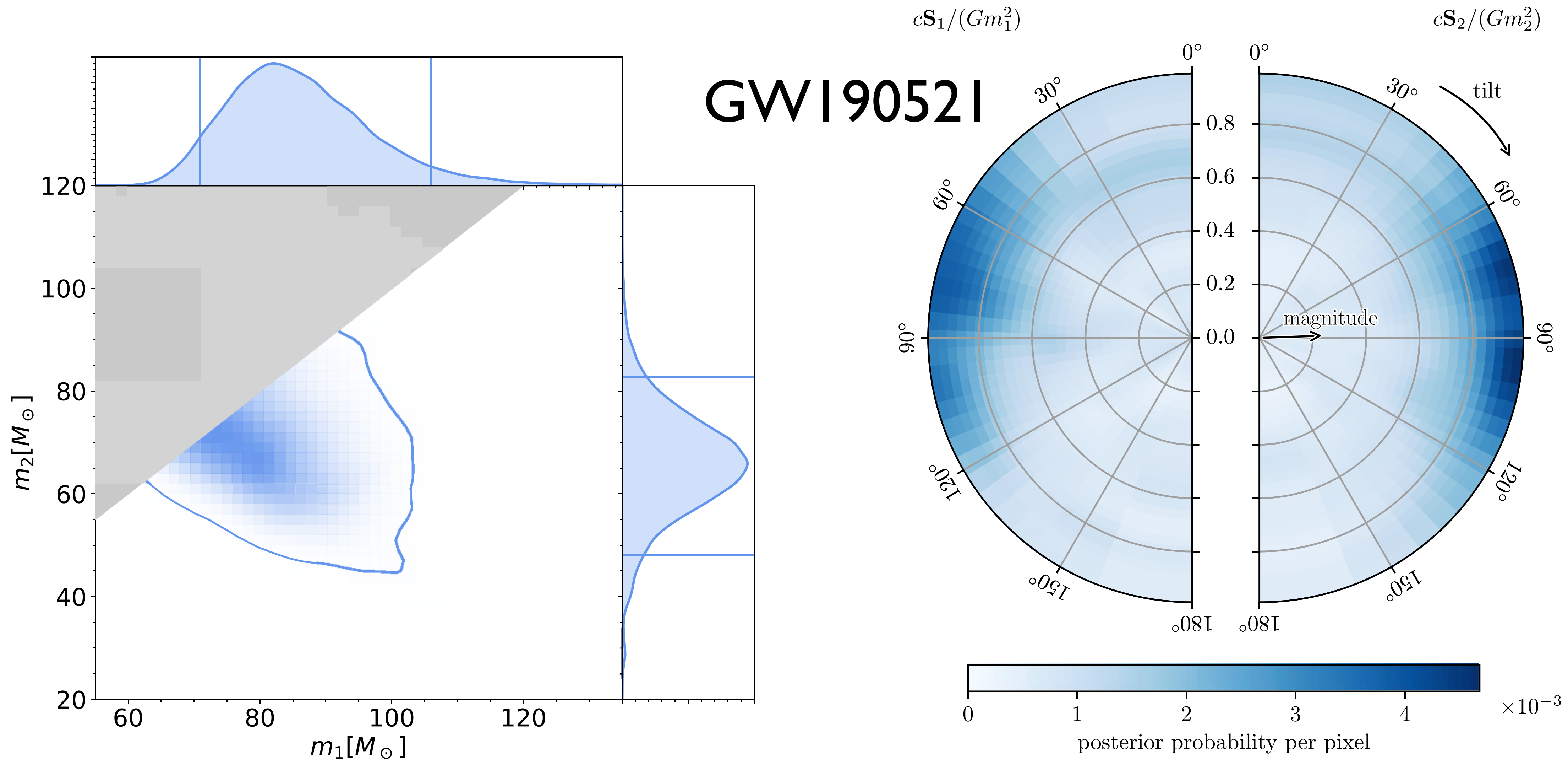
- Secondary information: amplitude in each detector

- With two detectors, time delays give a ring on the sky

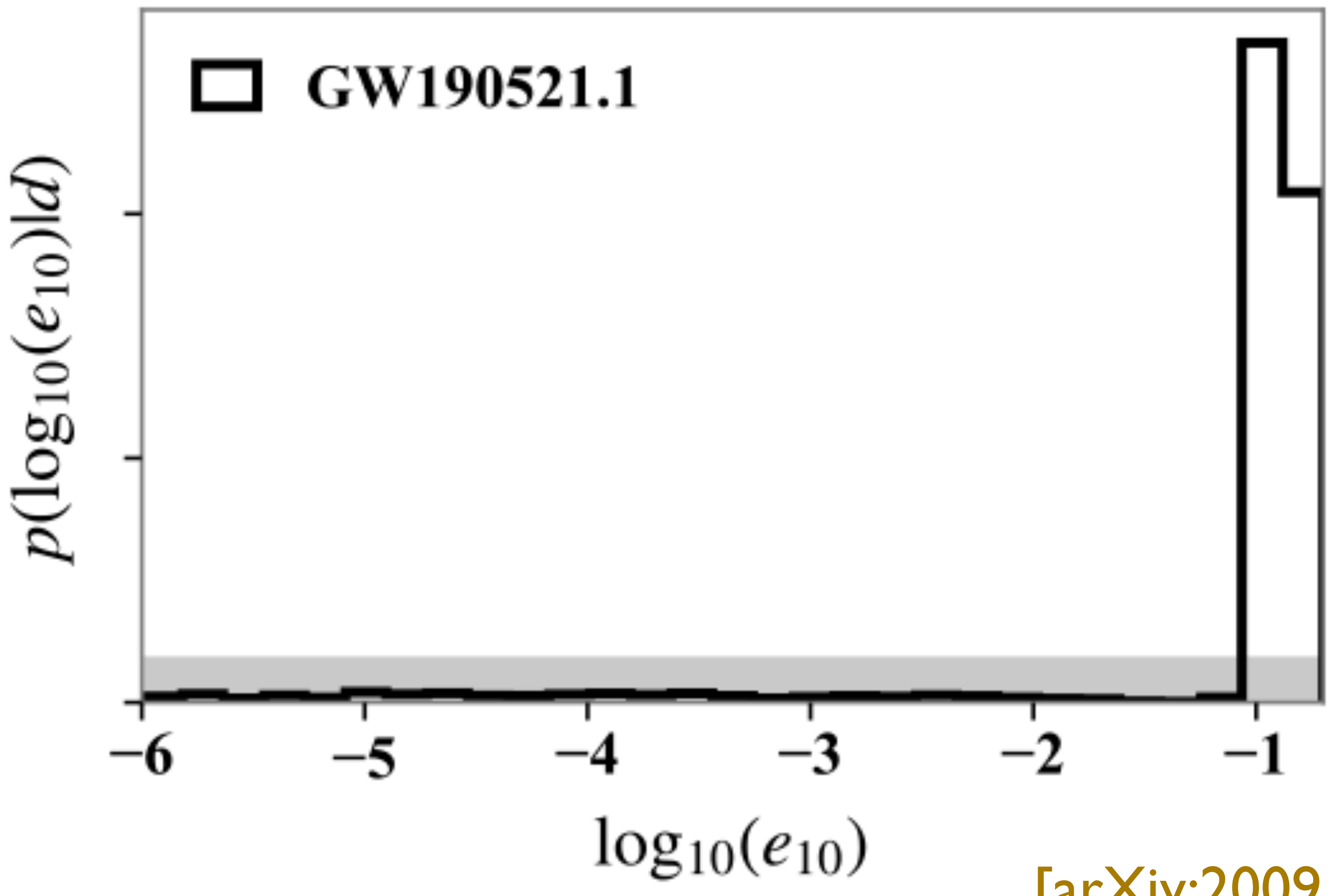
- Low-latency localization crucial even if approximate: Bayestar [arXiv:1508.03634]

PE challenges: systematics, exceptional events



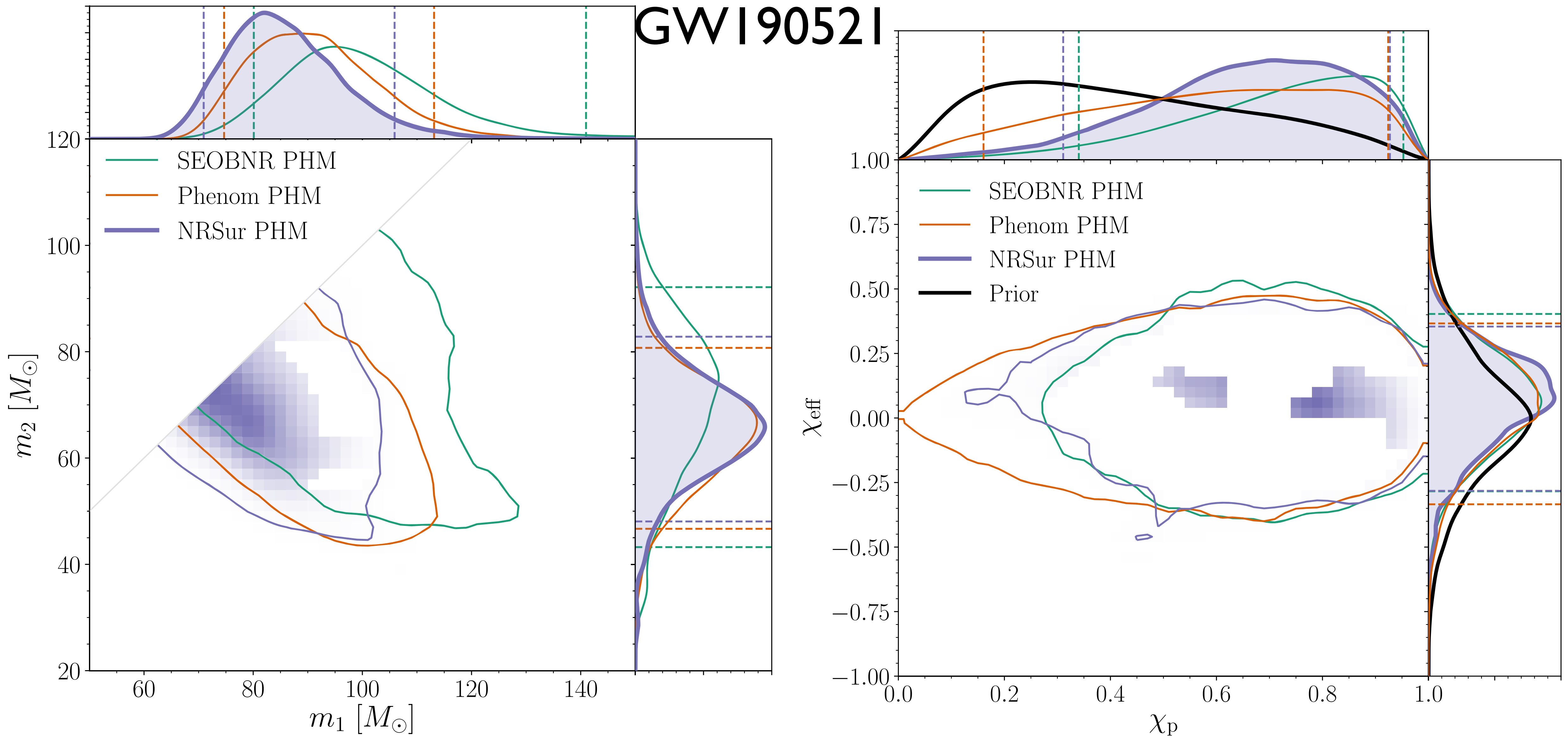


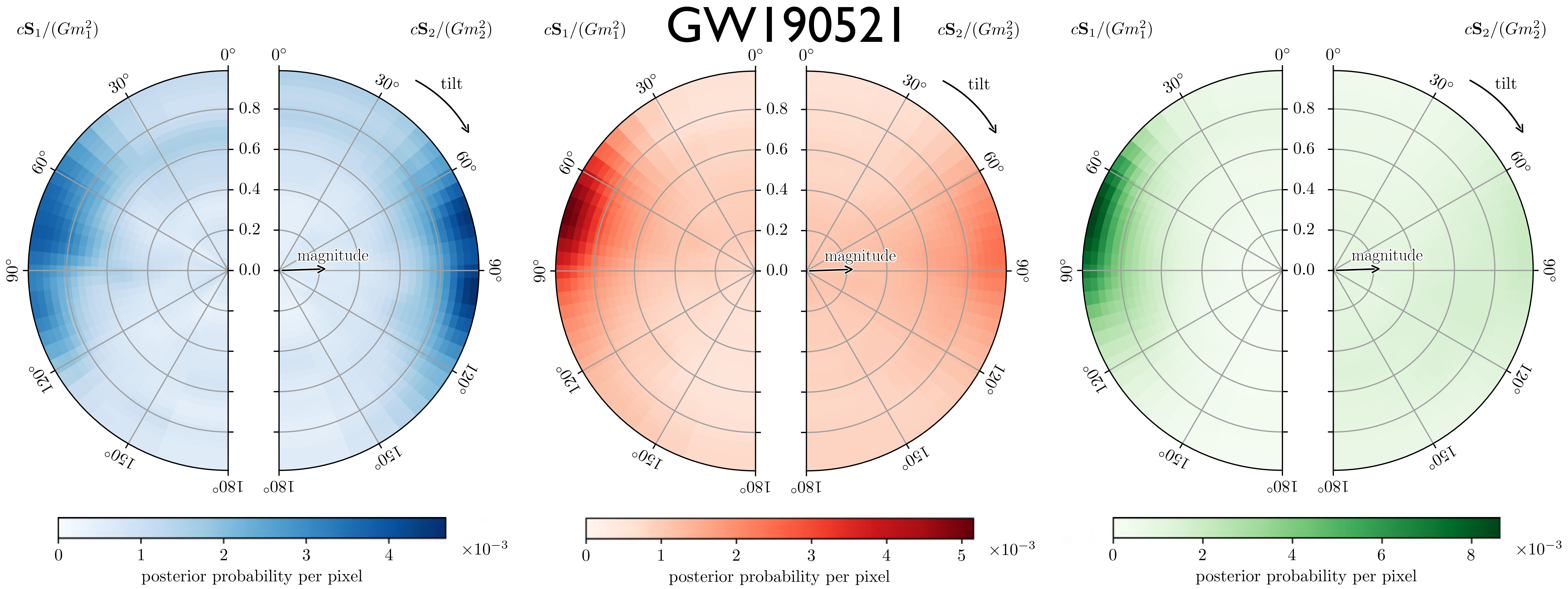
- GW190521 was an exceptionally massive BBH merger
- Surprising properties: in-plane spin ?
- Suggestion (to be confirmed) that eccentricity might be important

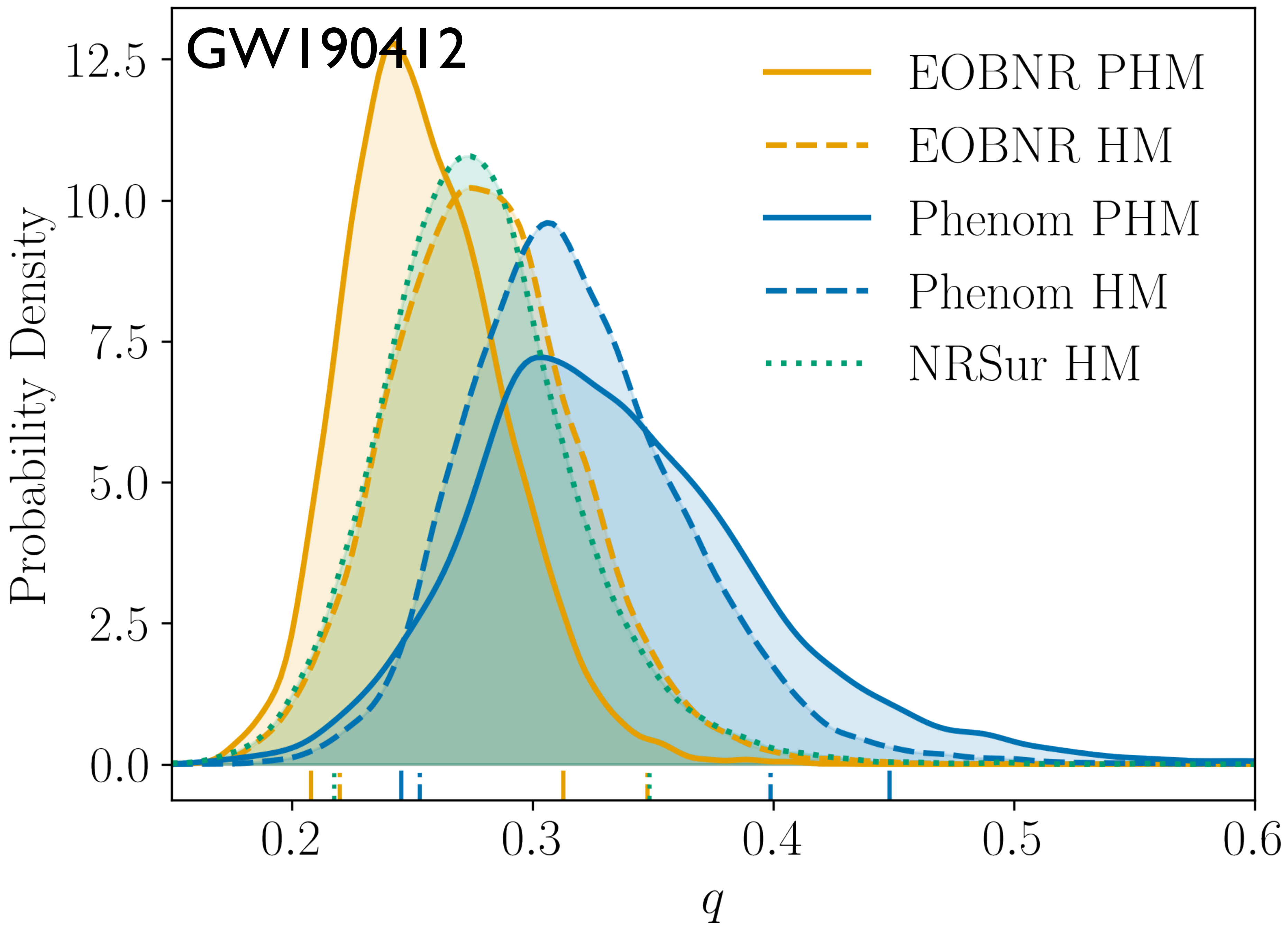


[arXiv:2009.04771]
[arXiv:2009.05641]

PE challenges: systematics, exceptional events



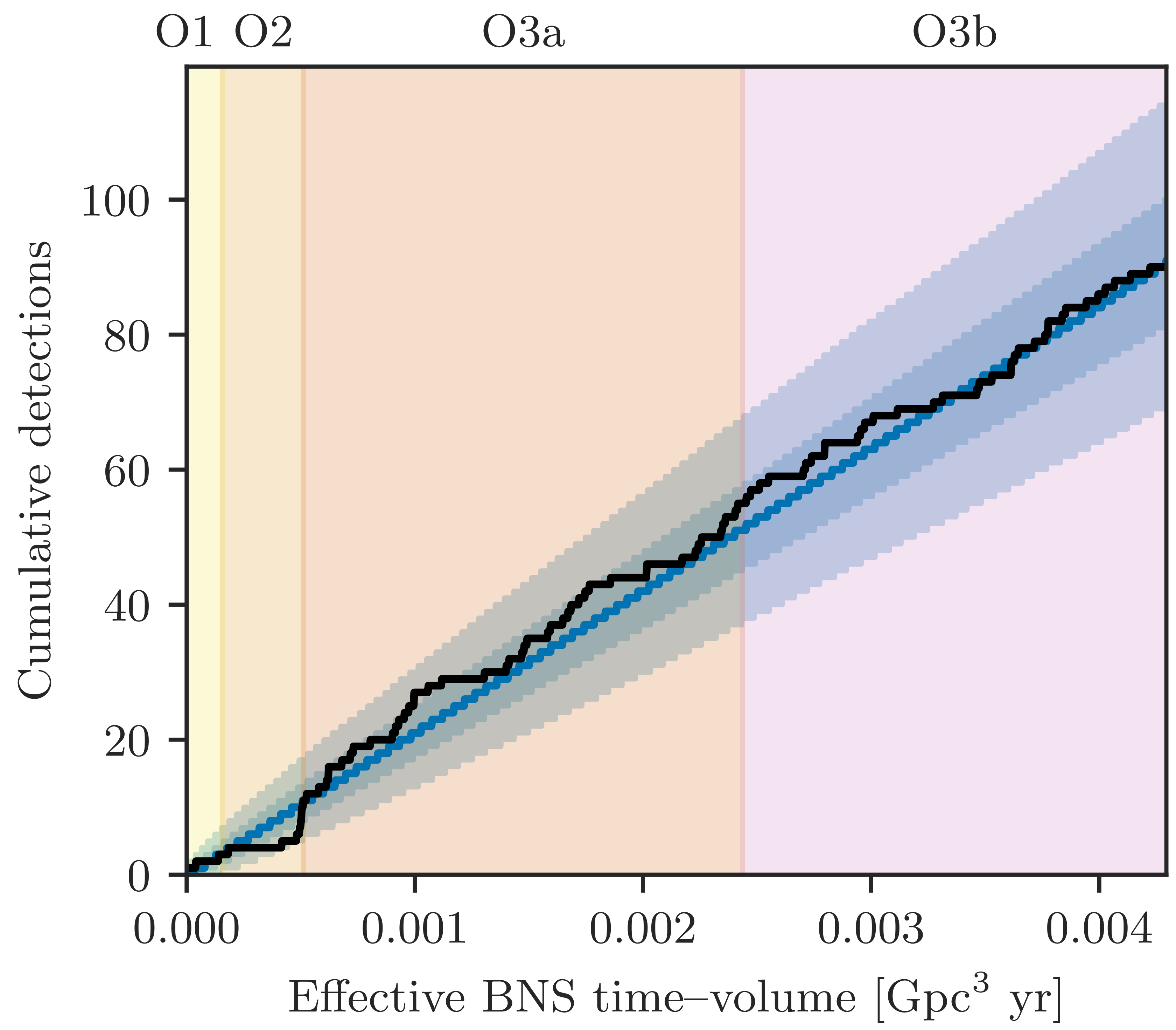


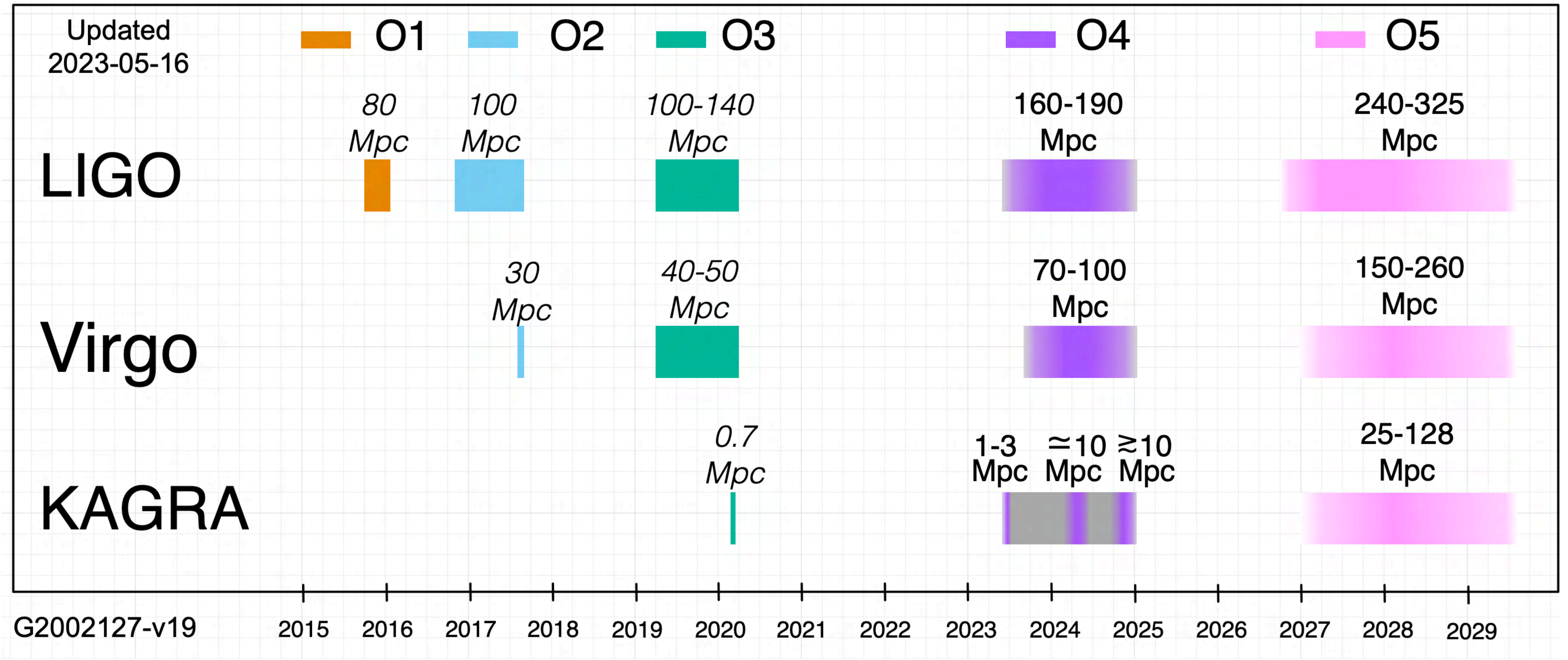


- Examples with disagreement between waveform models…

- Need improvement in models ?

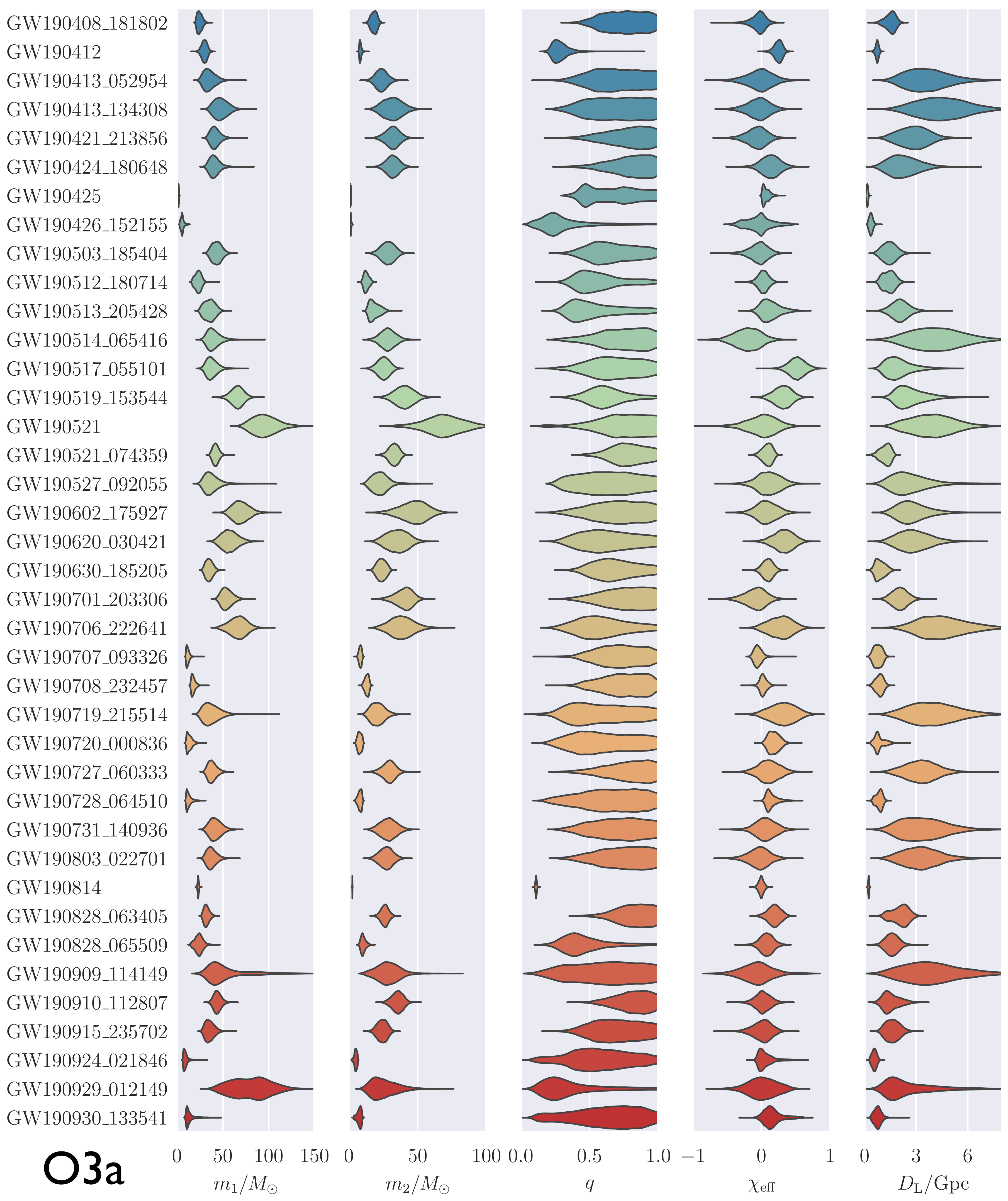
PE challenges: number of events



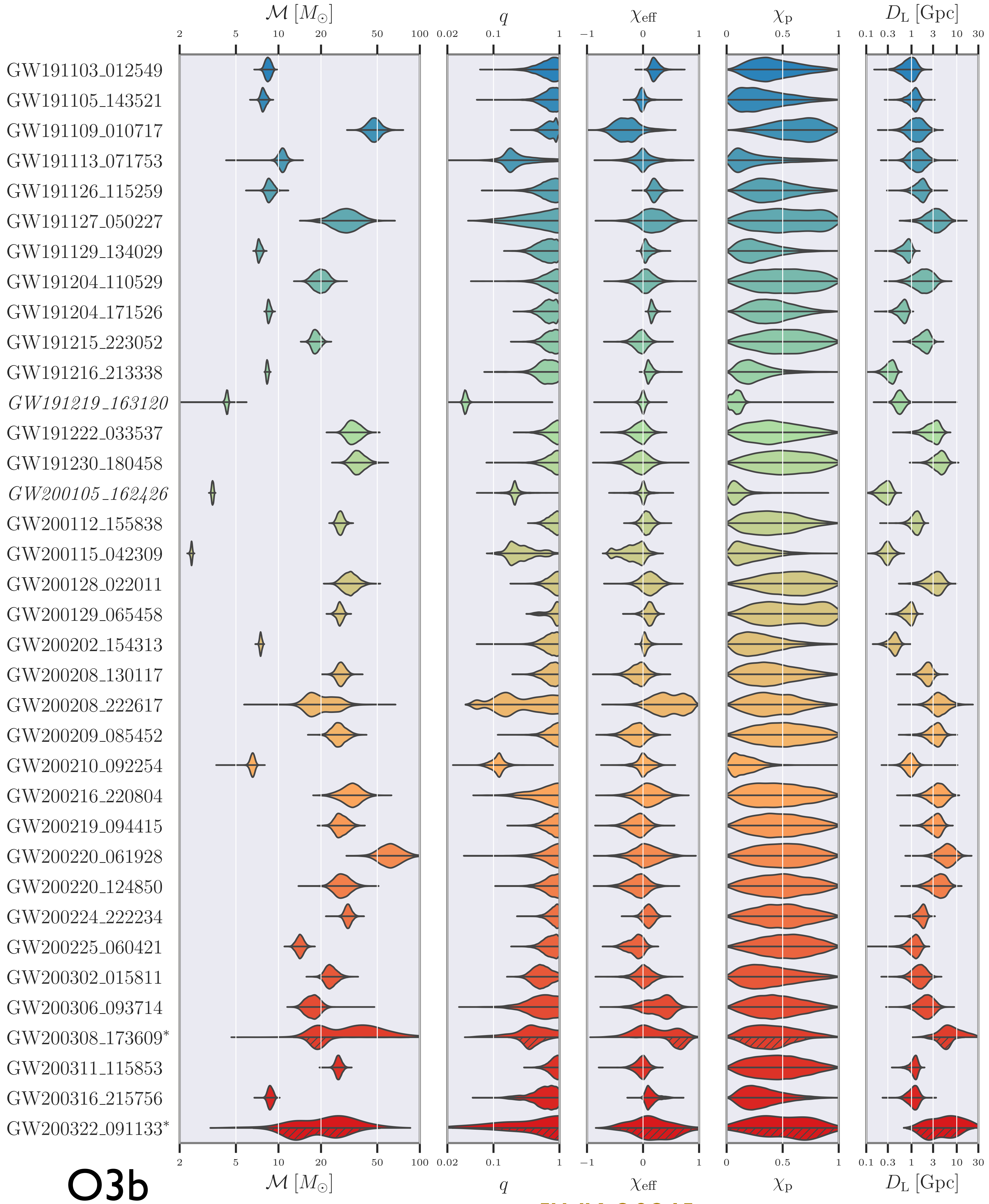


- Expecting higher event rates: R~d^3
- High SNRs for exceptional events: bigger challenge
- Computational challenge ! Automatization required

PE results from LIGO/Virgo: catalog



O3a



O3b

[LVK 2021]

Hierarchical Bayesian inference

Infer hyperparameters affecting the whole population (population model, cosmology, modified gravity)

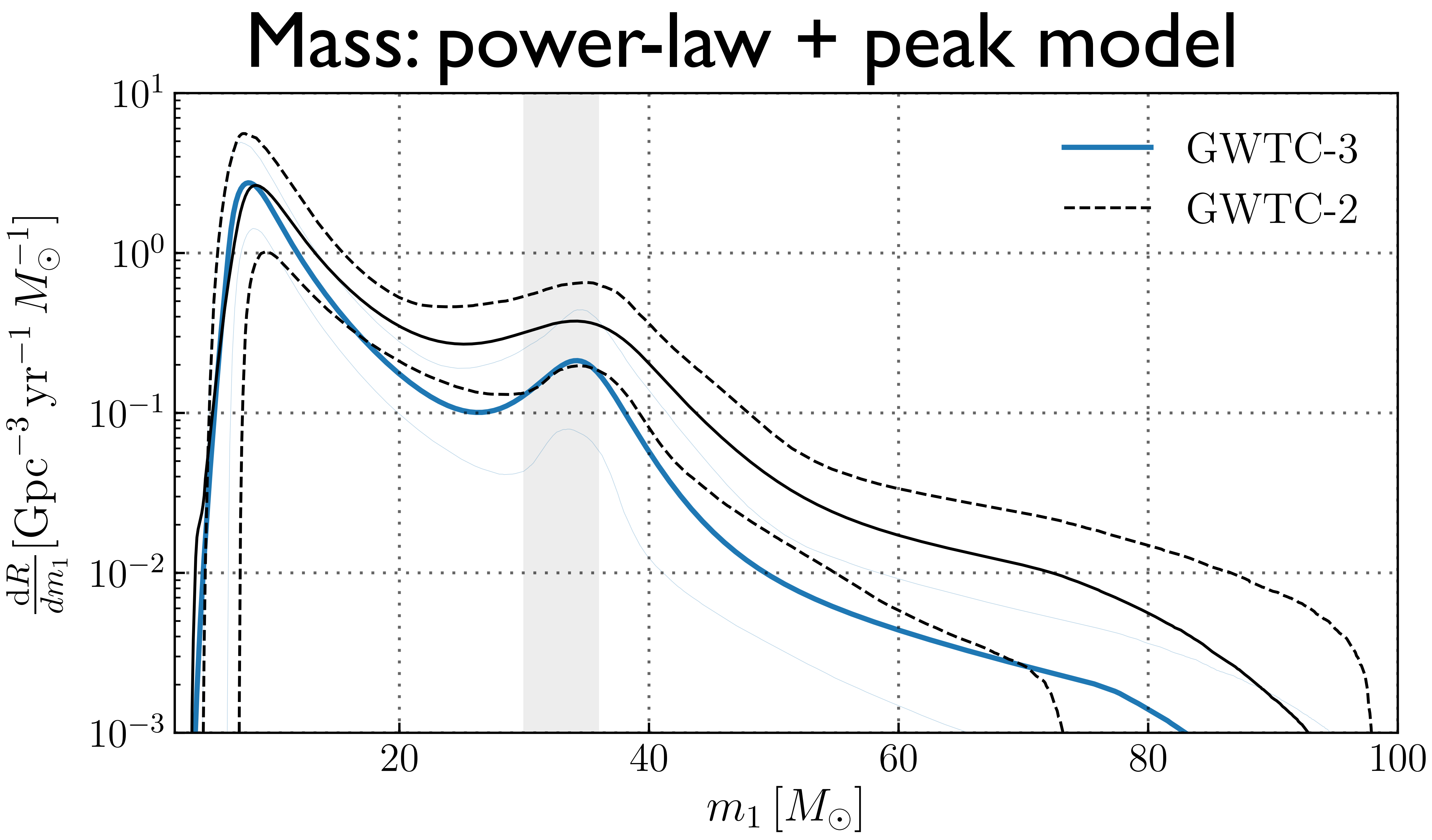
$$p(\Lambda|\{d\}) \propto p(\Lambda) \prod_{i=1}^{N_{\rm GW}} \frac{1}{\xi(\Lambda)} \int d\theta \, \mathcal{L}(d_i|\theta, \Lambda) p(\theta|\Lambda)$$

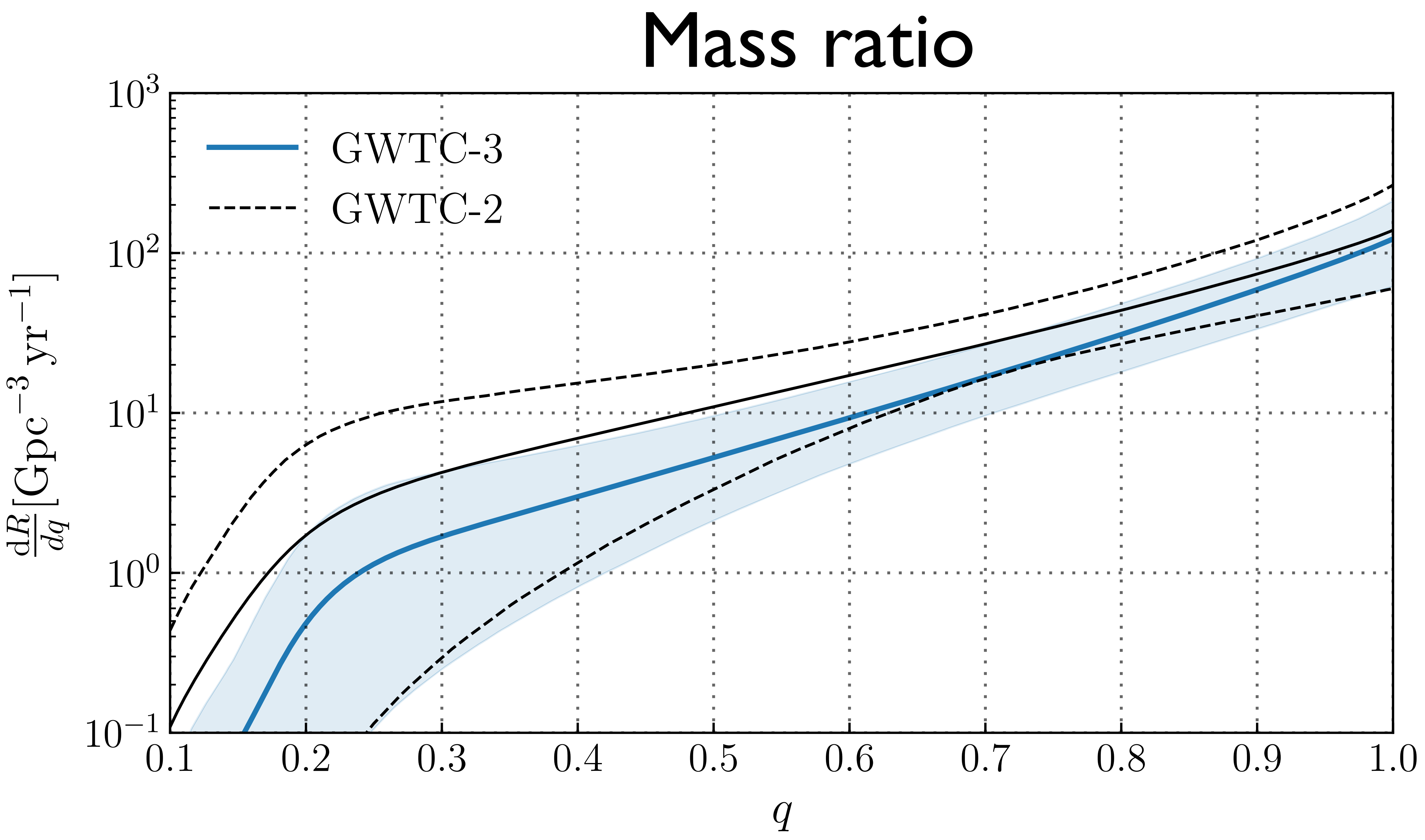
Selection effect: Malmquist bias, louder events more likely to be detected

Represented by factor $$\xi(\Lambda) = \int d\theta p_{\rm det}(\theta, \Lambda)$$

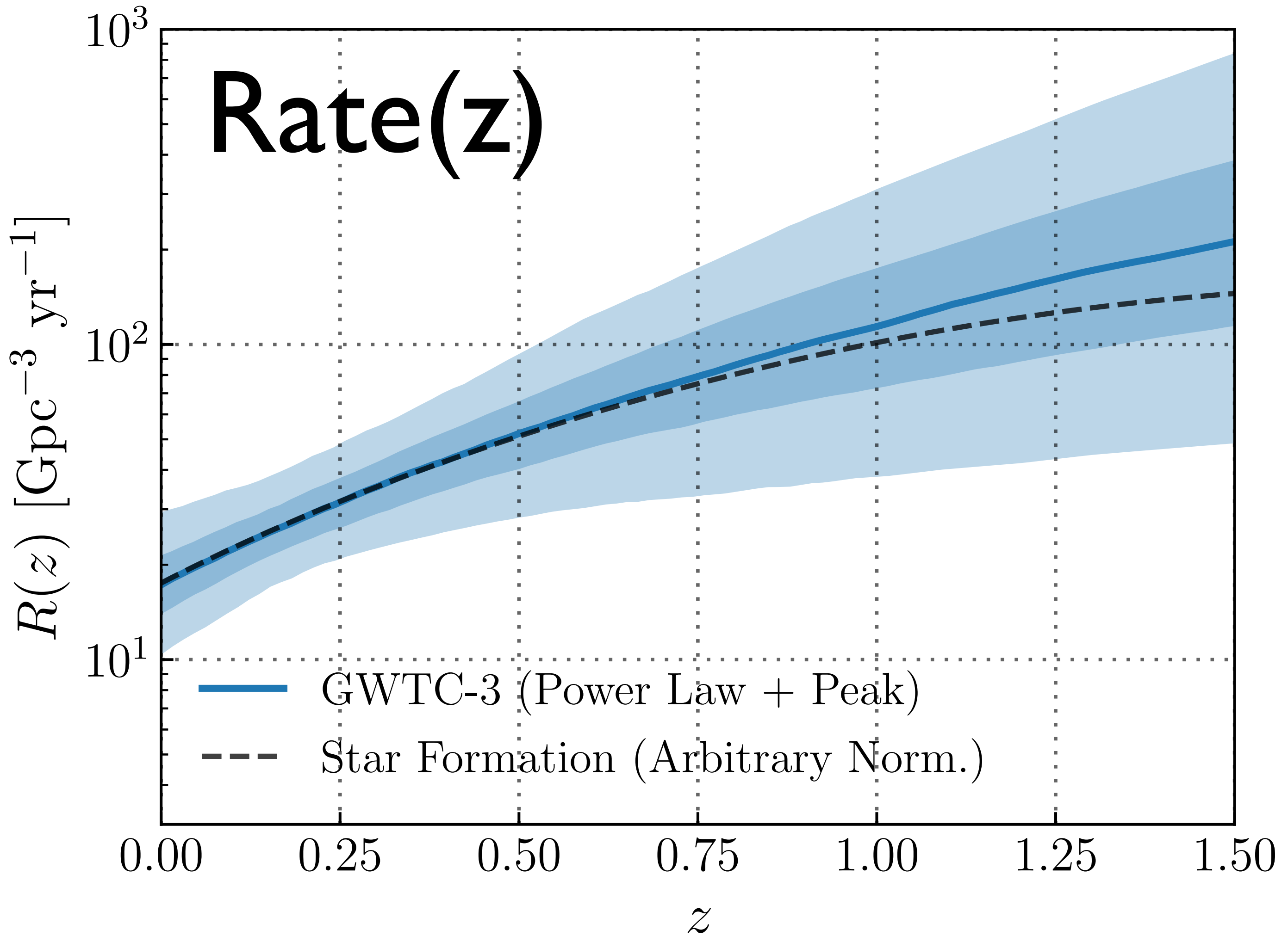
$$p_{\rm det}(\theta, \Lambda) = \int_{x > \rm thres.} dx \, \mathcal{L}(x|\theta, \Lambda)$$

Results from LIGO/Virgo: population



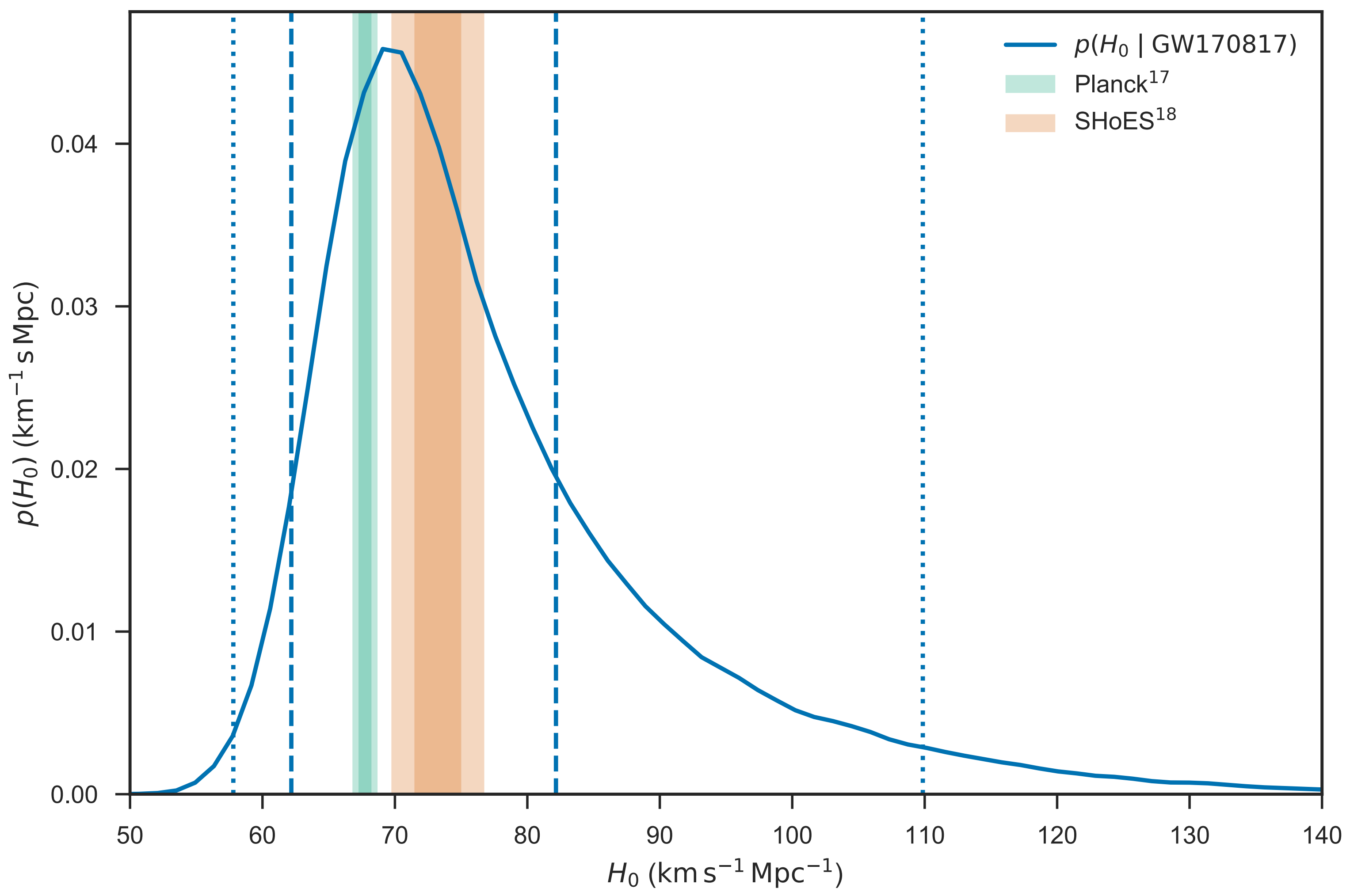


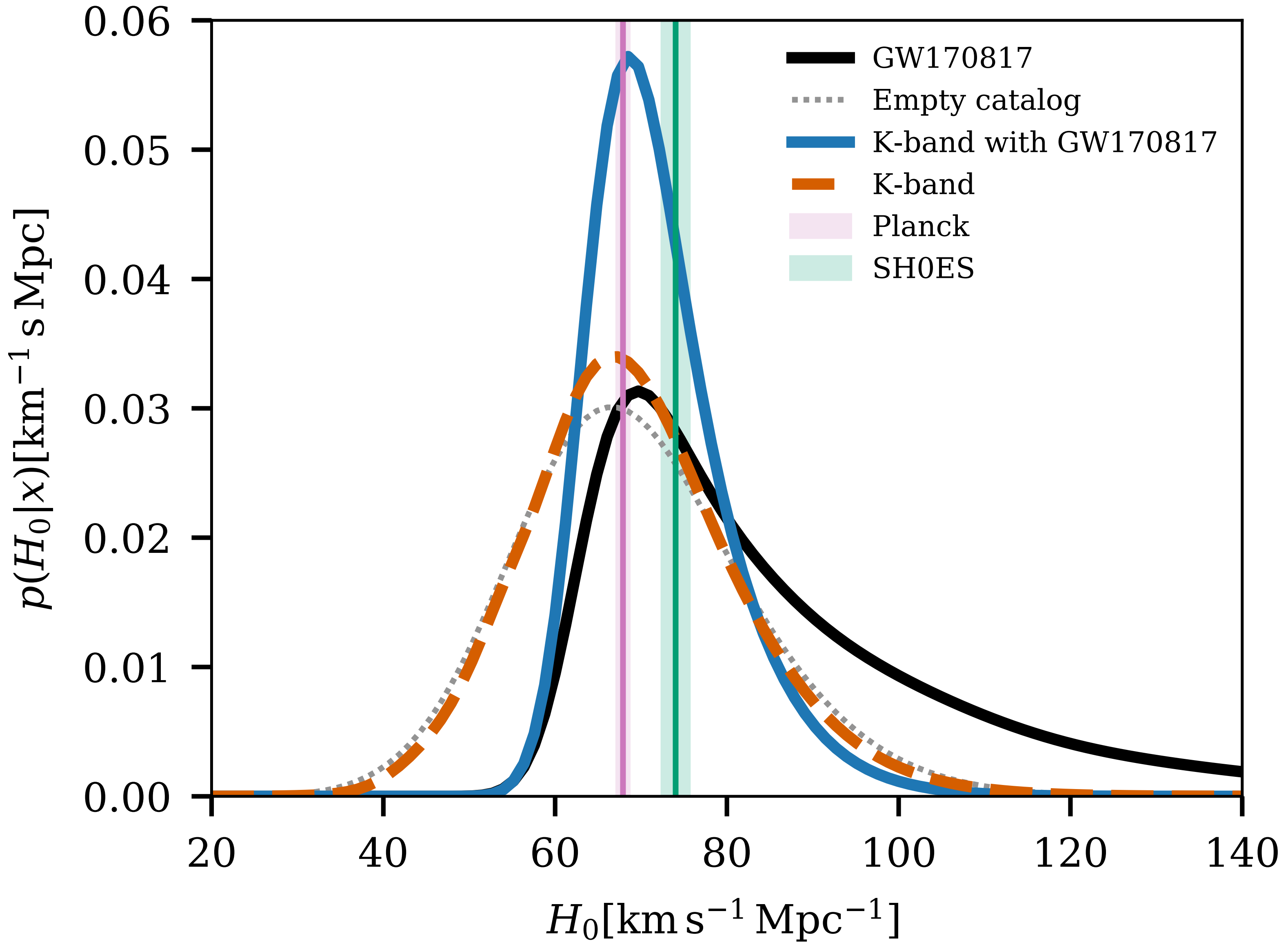
- Introduce parametrized population models
- Hierarchical parameter estimation to estimate parameters (with uncertainties)
- Represent posterior predictive distributions

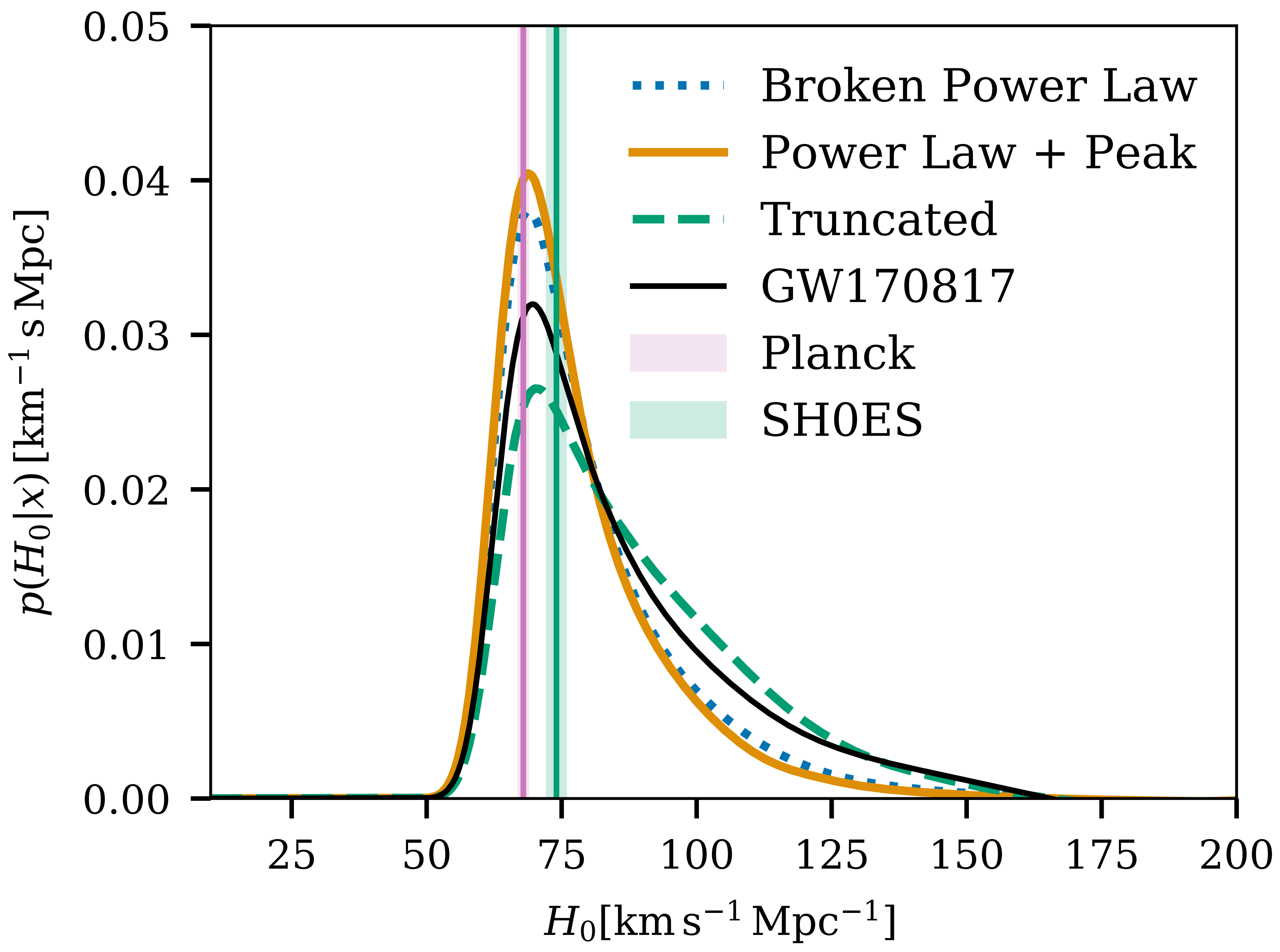


Results from LIGO/Virgo: cosmology

GW measure luminosity distance; if able to get redshift information, constrain dL(z) and cosmological parameters







- Cosmology with an EM counterpart: GW170817
- Dark sirens: cross-correlation with galaxy catalogs
- Spectral sirens: learn source-frame mass from feature in mass distribution

Outline

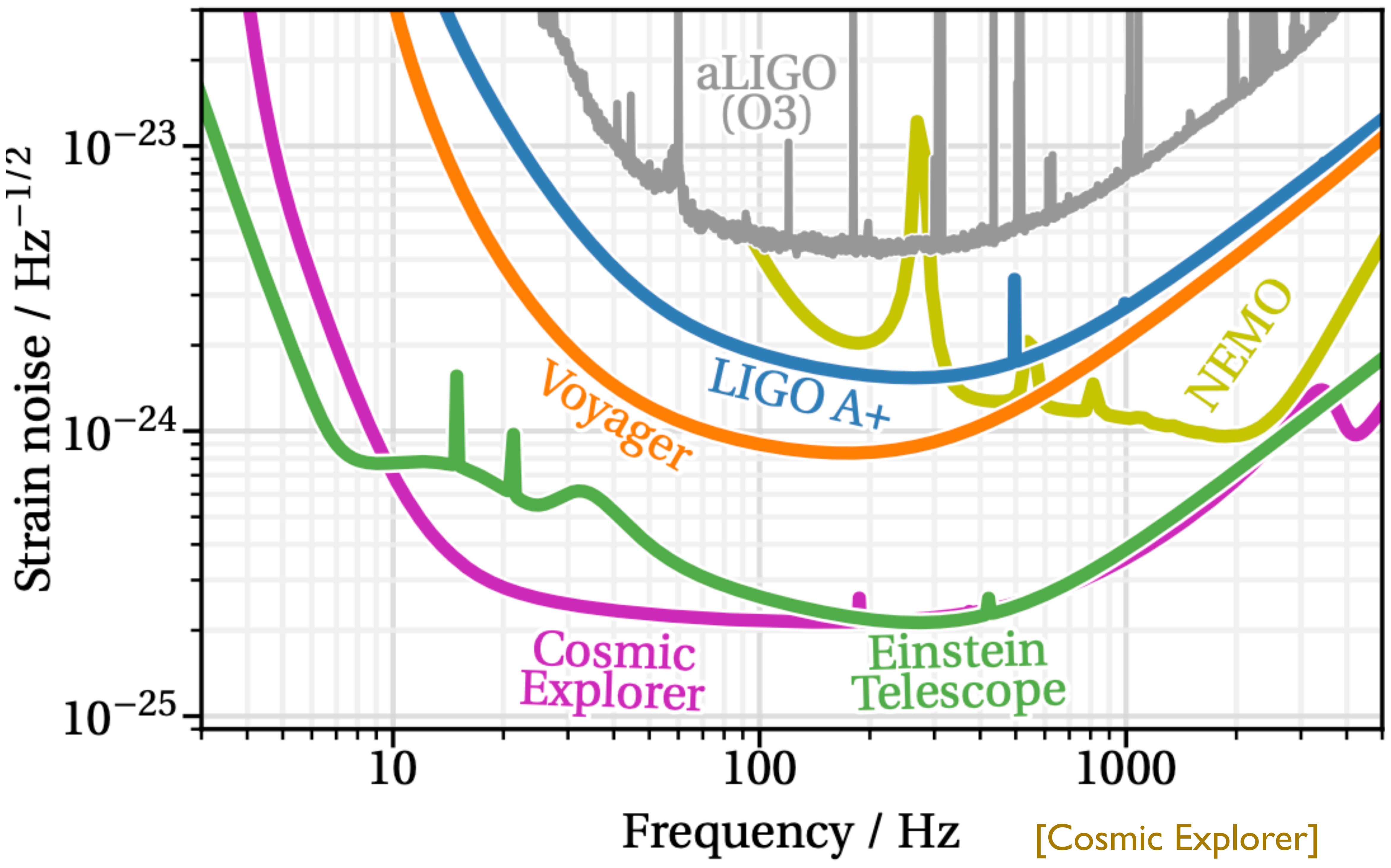
Part I

- GW signals: the basics
- Noise as a stochastic process
- Introducing matched filtering
- Towards real CBC searches
- Other signals: continuous waves, stochastic backgrounds

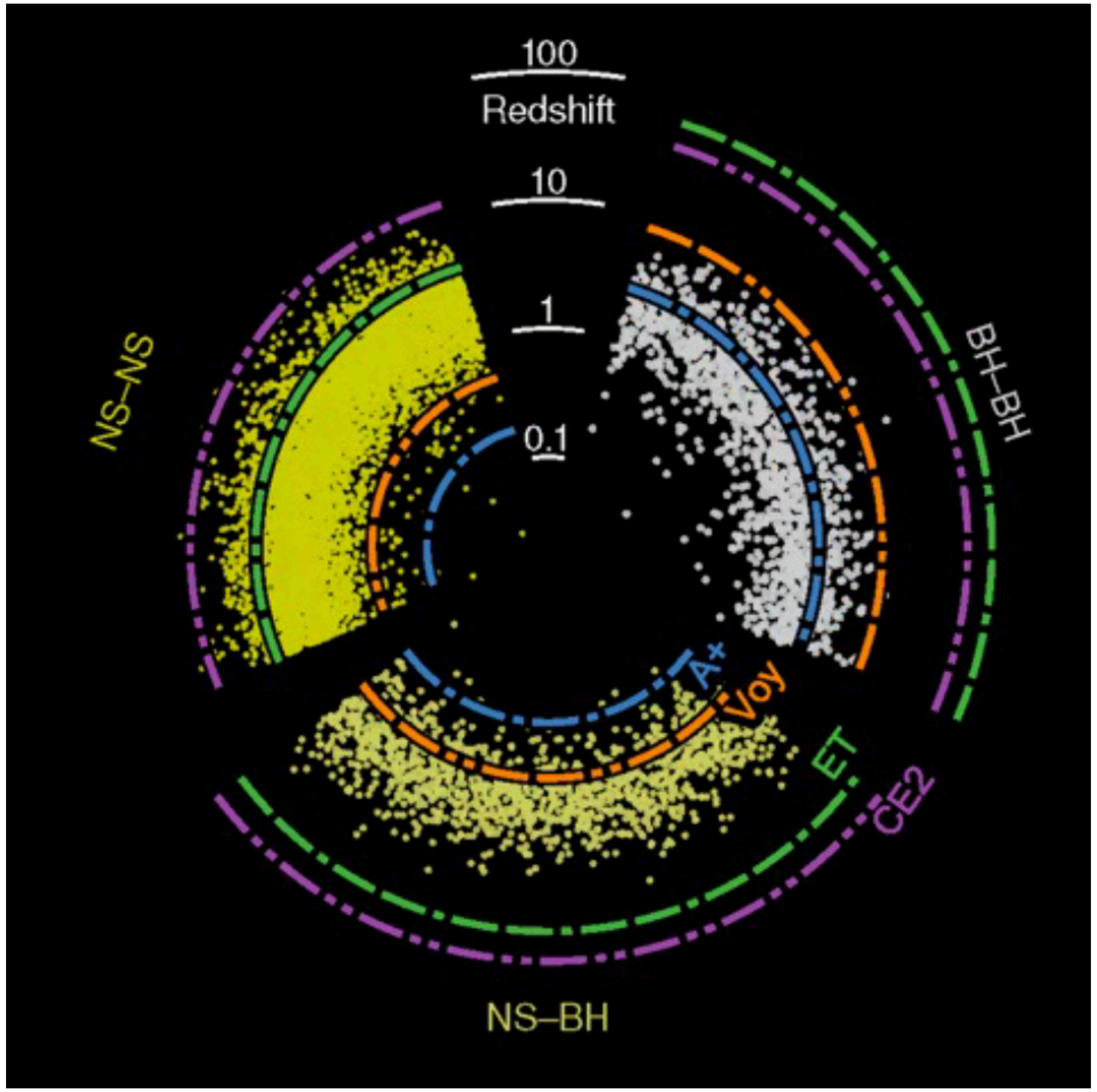
Part II

- Bayesian parameter estimation basics, likelihood
- Parameter space and waveforms
- Fisher matrix approach
- Metropolis-Hastings MCMC, Parallel tempering and example PE
- PE toolbox
- PE results from LVK
- **Future detectors and their challenges**

3G detectors



[Cosmic Explorer]



[Cosmic Explorer]

Events/yr (low-median-high):

- BBH: 60k-90k-150k
- BNS: 300k-1000k-3000k

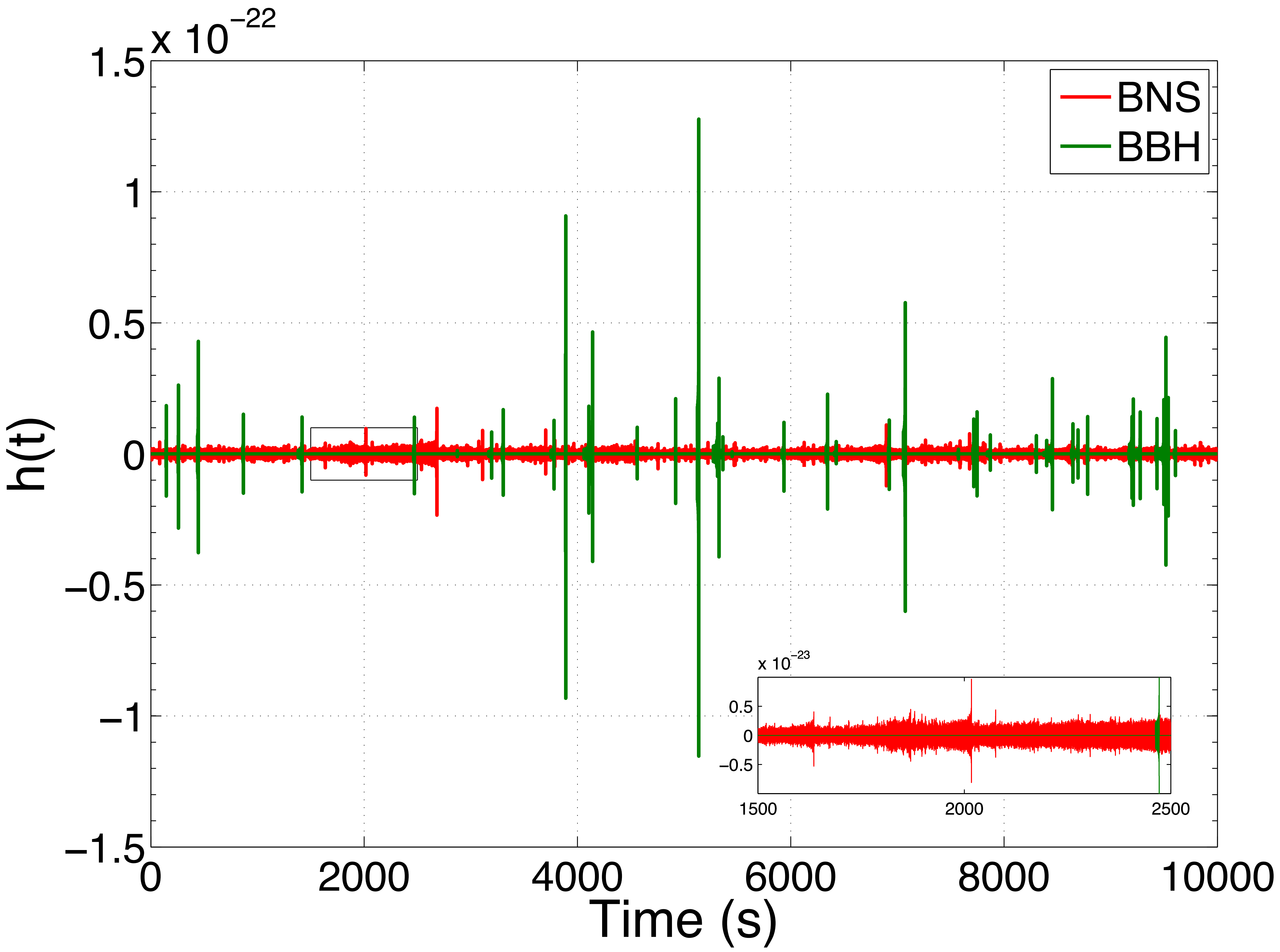
Detections (2 CE+1ET):

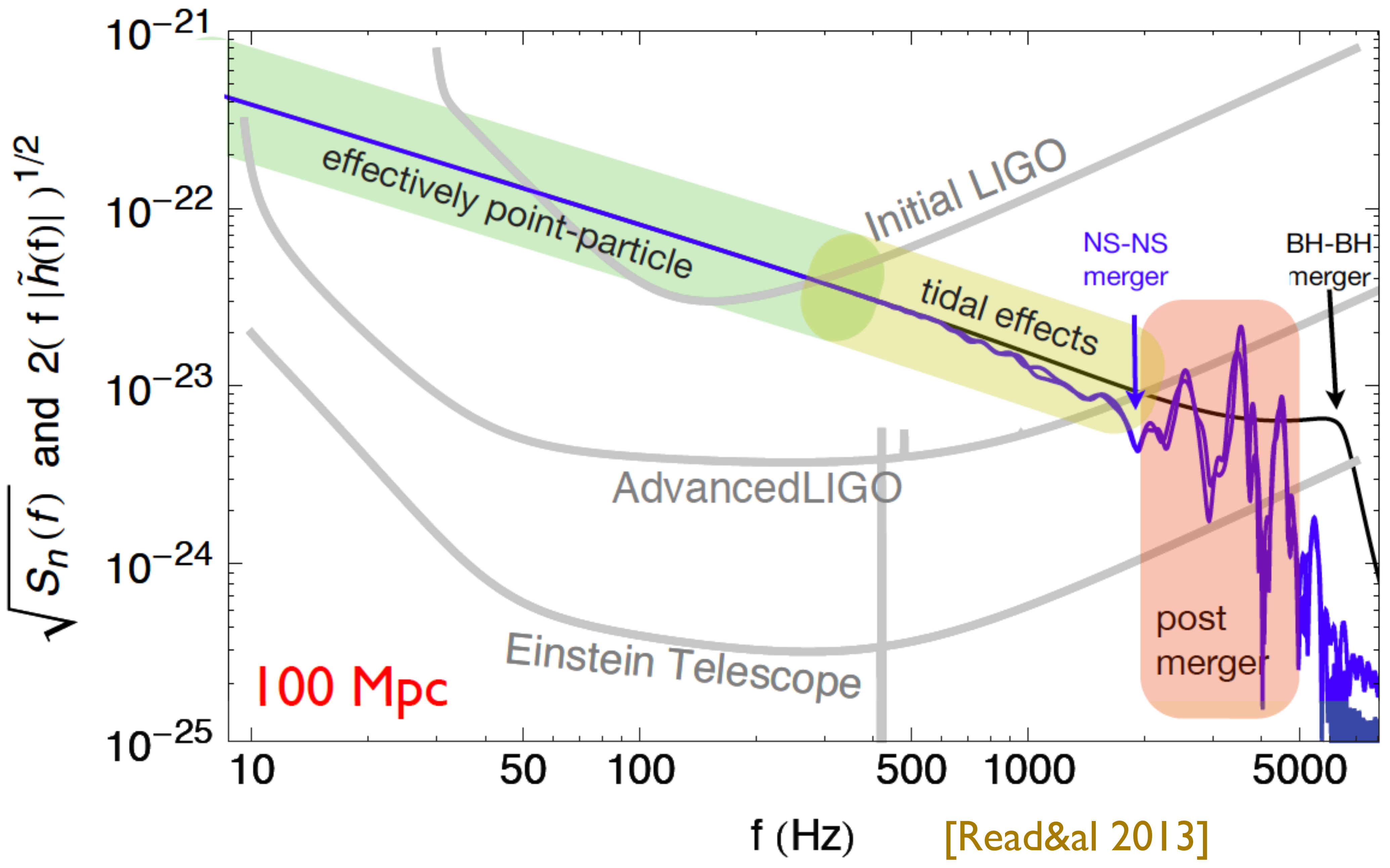
- BBH: 93%
- BNS: 35%

[arXiv:2102.07544]

Computational challenge !

3G detectors

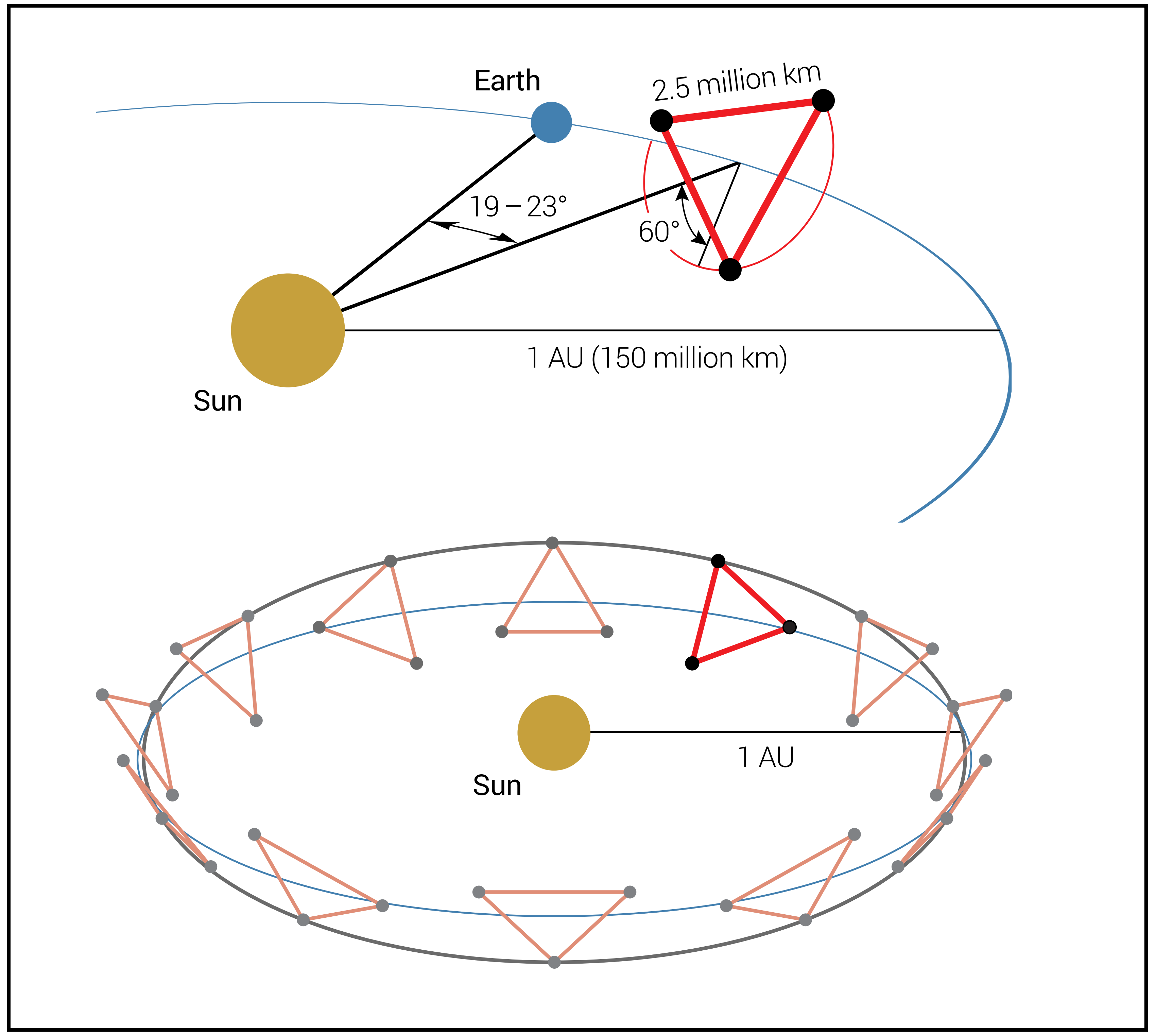




- Post-merger BNS signal
- Analysis of very long-lived BNS

- Popcorn nature of combined signals
- Superposition problem

LISA instrument



Doppler delay from orbit, change in orientation

Analogous to 2 LIGO in motion at low frequencies only

From spacecraft s to spacecraft r through link s: $y = \Delta\nu/\nu$

$$y_{slr} = \frac{1}{2}\frac{1}{1-\hat{k}\cdot n_l} n_l \cdot (h(t_s) - h(t_r)) \cdot n_l$$

Response time and frequency-dependent:

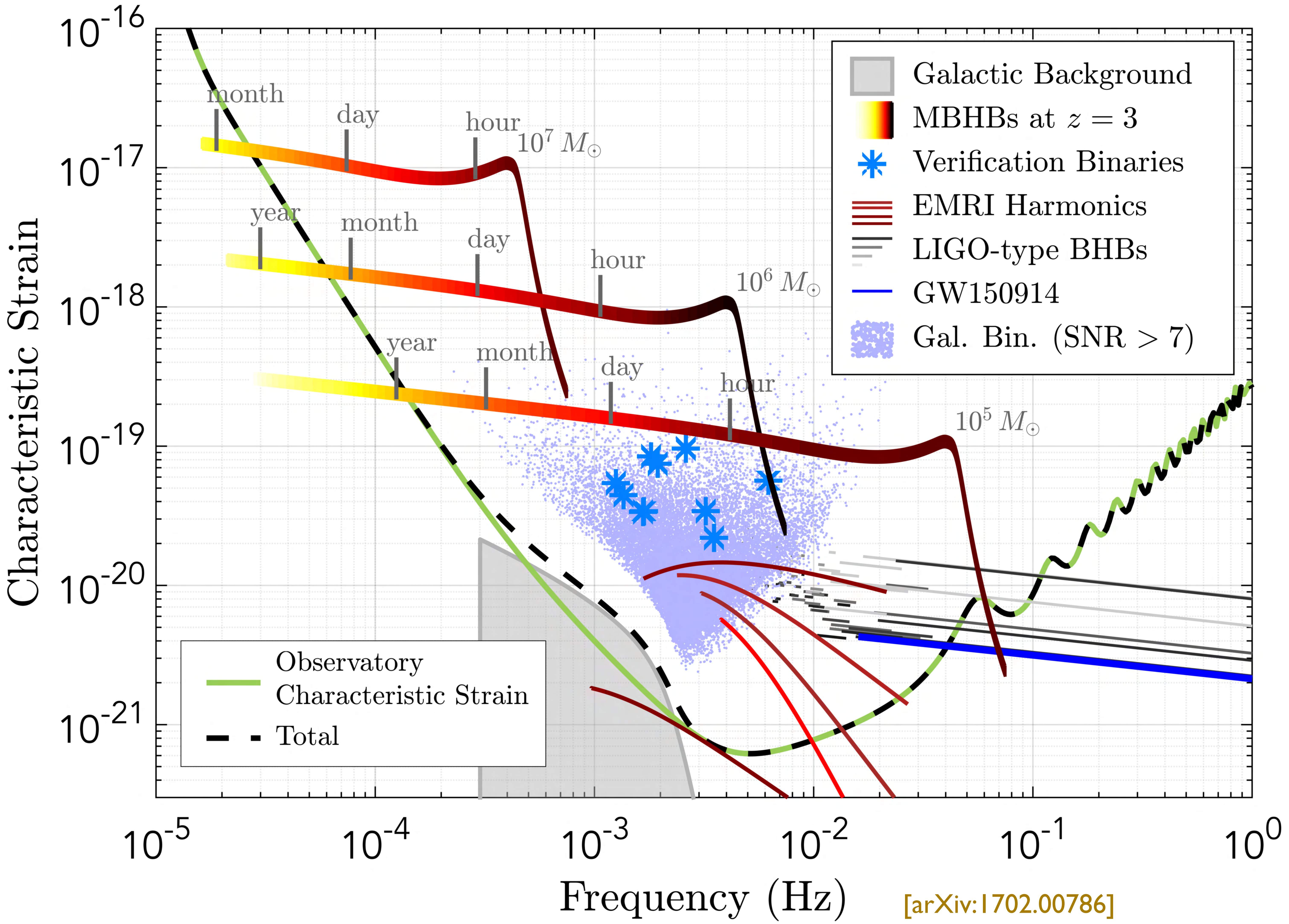
$$\mathcal{T}_{slr} = \frac{i\pi f L}{2} \mathrm{sinc}\left[\pi f L \left(1 - k \cdot n_l\right)\right] \exp\left[i\pi f \left(L + k \cdot \left(p_r + p_s\right)\right)\right] n_l \cdot P \cdot n_l(t_f)$$

+ Time-delay interferometry (TDI)

$$X_1^{\rm GW} = \underbrace{\left[(y_{31}^{\rm GW} + y_{13,2}^{\rm GW}) + (y_{21}^{\rm GW} + y_{12,3}^{\rm GW})_{,22} - (y_{21}^{\rm GW} + y_{12,3}^{\rm GW}) - (y_{31}^{\rm GW} + y_{13,2}^{\rm GW})_{,33}\right]}_{X^{\rm GW}(t)}$$
$$- \underbrace{\left[(y_{31}^{\rm GW} + y_{13,2}^{\rm GW}) + (y_{21}^{\rm GW} + y_{12,3}^{\rm GW})_{,22} - (y_{21}^{\rm GW} + y_{12,3}^{\rm GW}) - (y_{31}^{\rm GW} + y_{13,2}^{\rm GW})_{,33}\right]_{,2233}}_{X^{\rm GW}(t-2L_2-2L_3)\simeq X^{\rm GW}(t-4L)} .$$

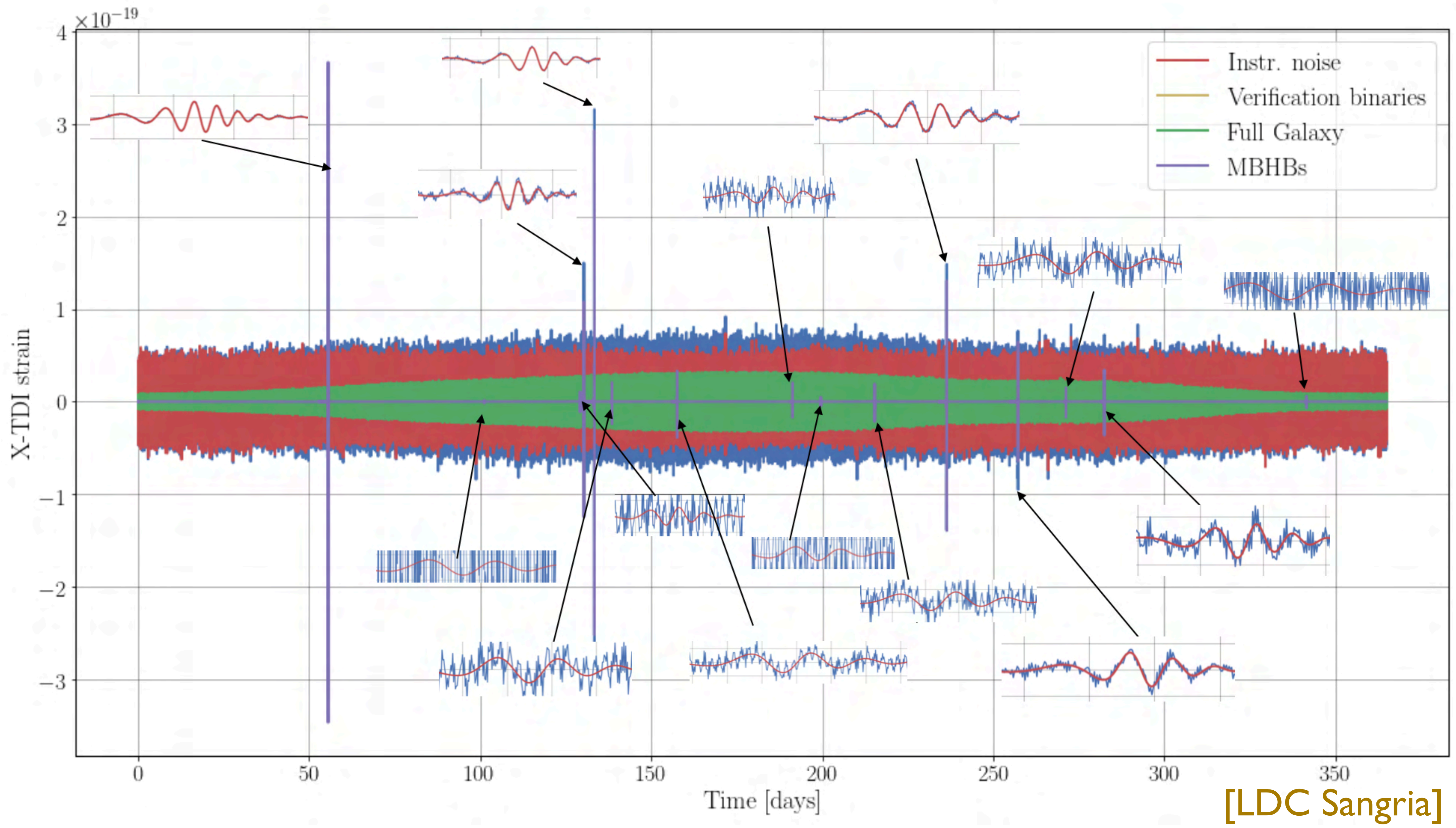
LISA sources

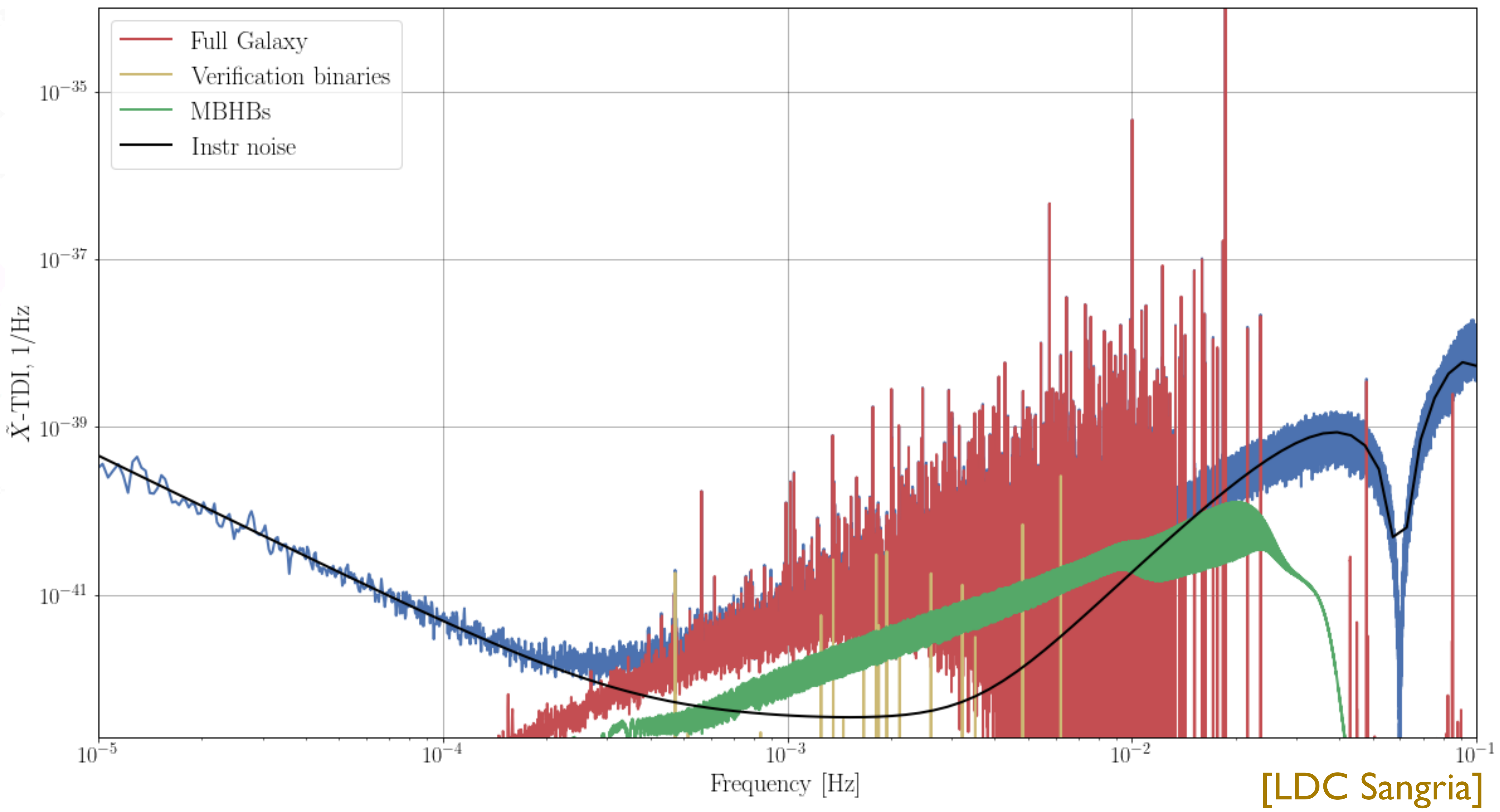
- Massive black holes binaries (MBHBs)
- Population of galactic binaries (DWD), confusion background
- Extreme mass ratio inspirals (EMRIs)
- Stellar-mass black hole binaries (SBHBs)
- Cosmological backgrounds ?



[arXiv:1702.00786]

LISA: data



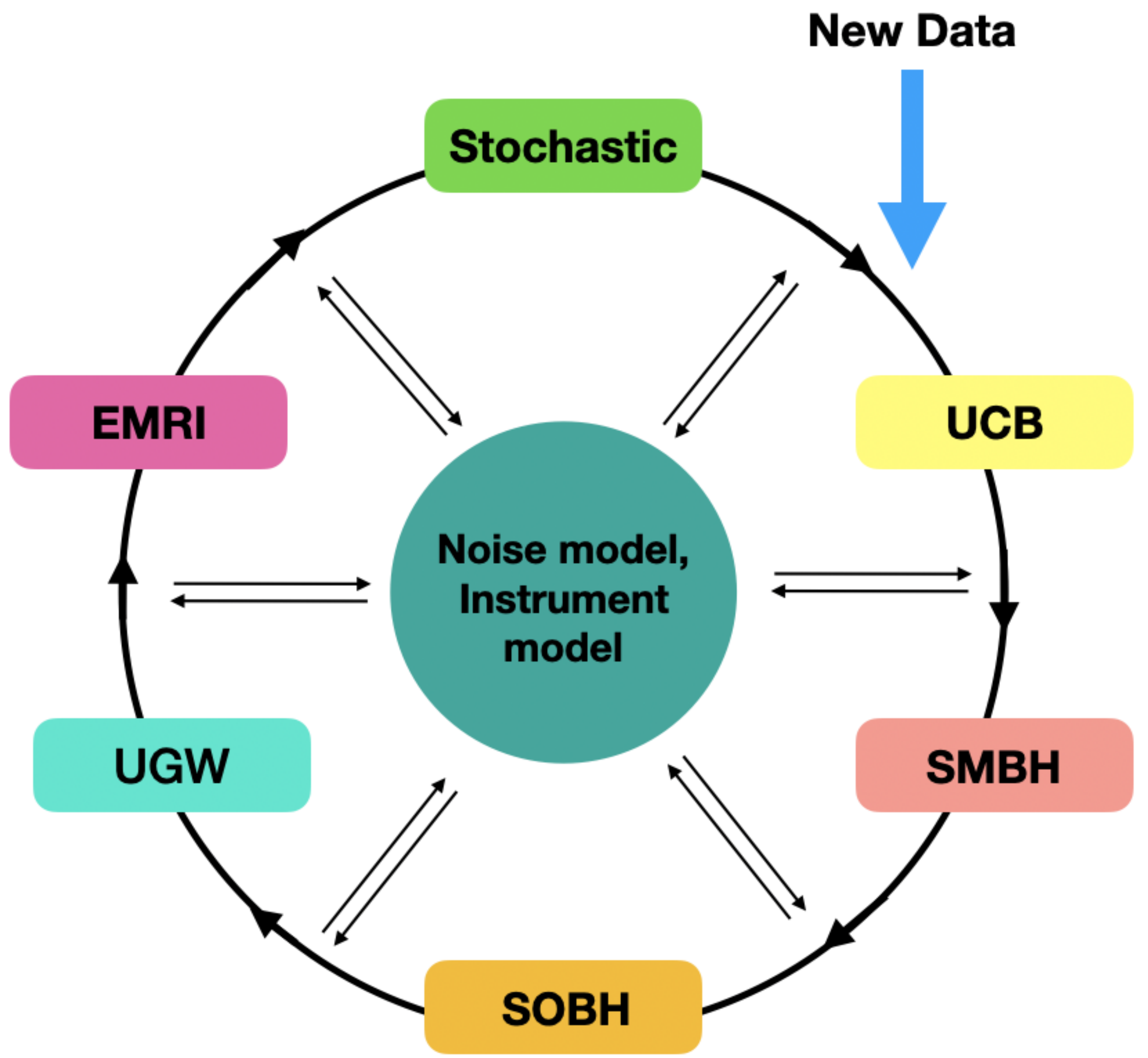


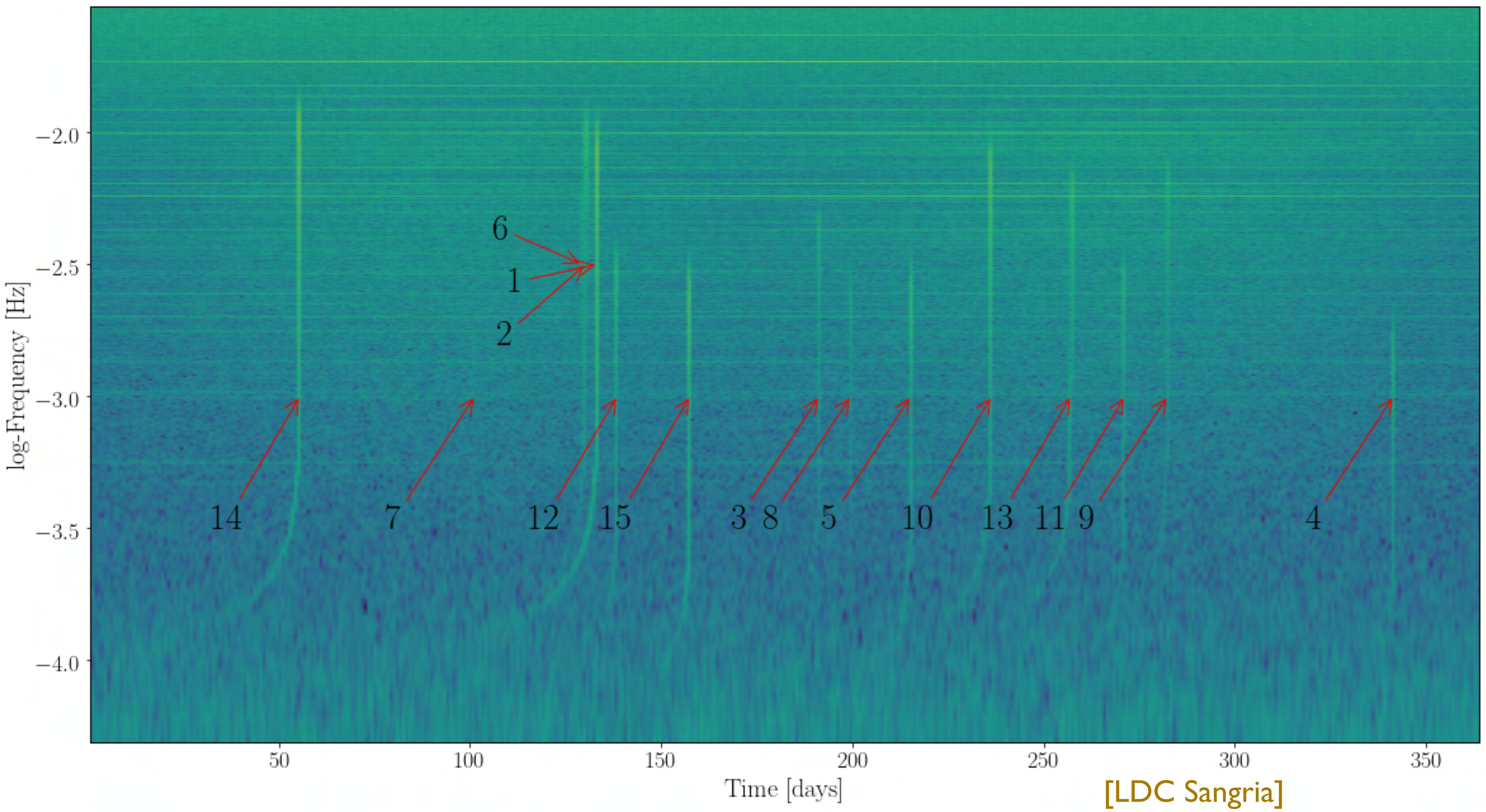
Superposition of sources !

LISA Data Challenges (LDC)

- 1st challenge (Radler): single class of sources
- 2nd challenge (Sangria): MBHBs, GBs, noise

LISA: global fit





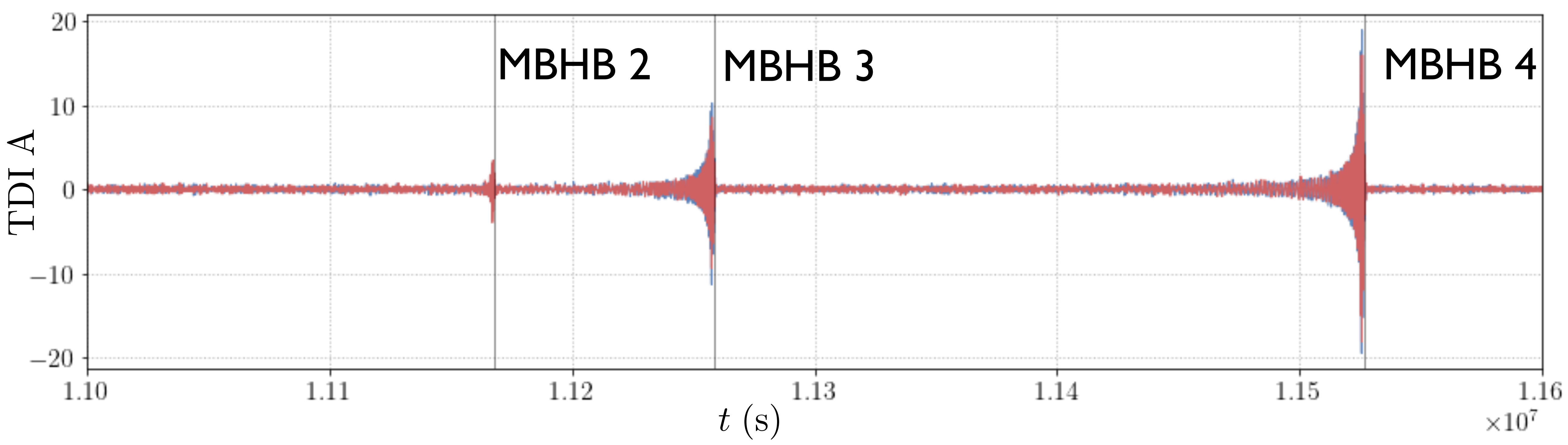
[LDC Sangria]

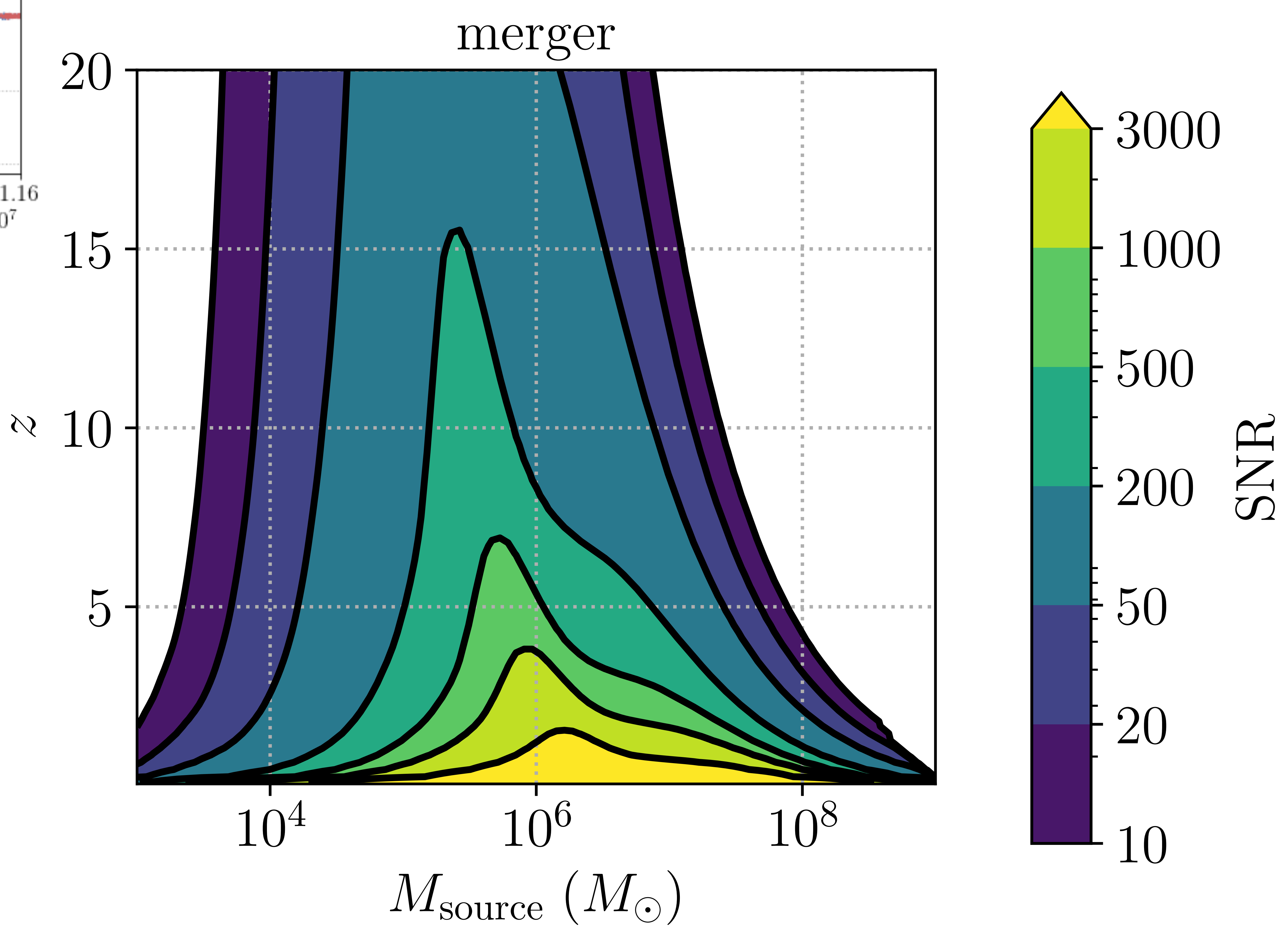
Global fit

- Raw dimensionality untractable
- Gibbs sampling across different source types
- Orthogonality between signals
- Noise has to be estimated as well (no signal-free segments)

LISA: challenges for MBHBs

Whitened, band-passed LDC data

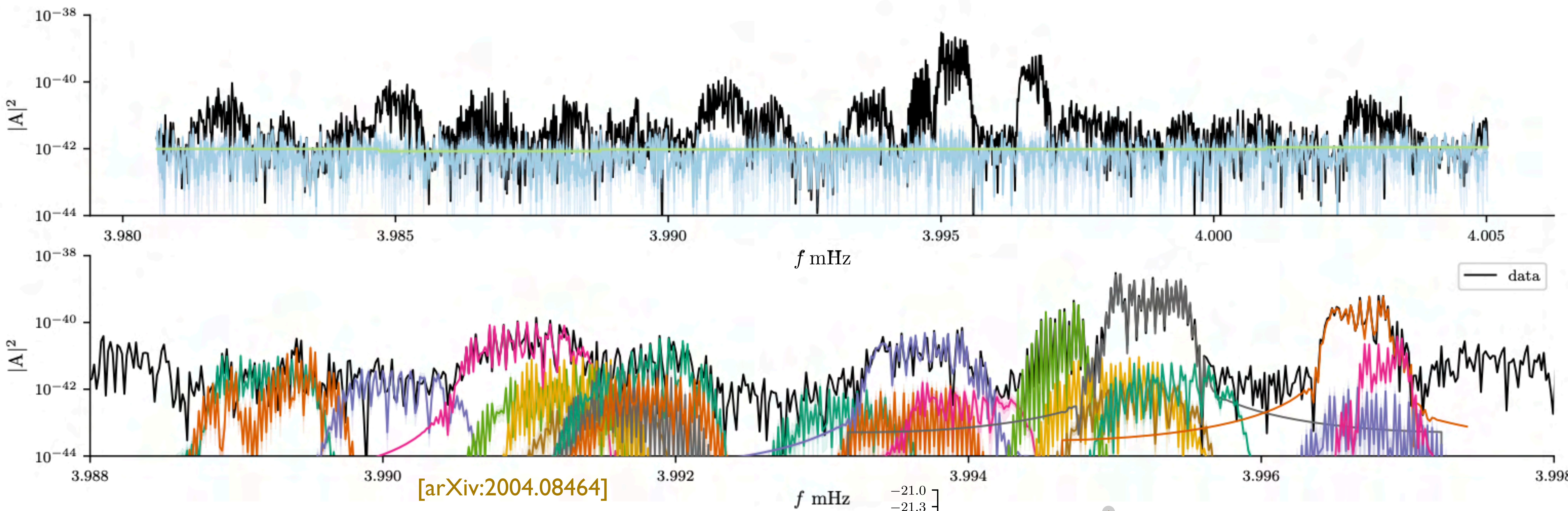




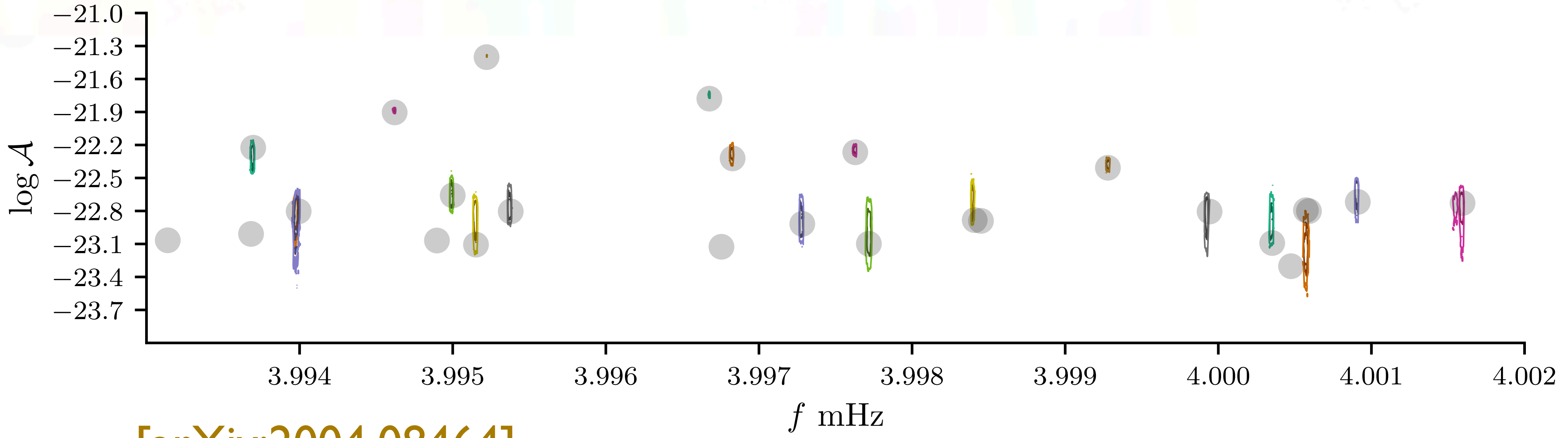
Unlike LVK BHBs:

- Very high SNRs, challenge for waveform modelling
- At high M, merger-dominated signals
- High spins ? Strong precession ?
- Eccentricities might be large
- Effect of superposition with other signals ?

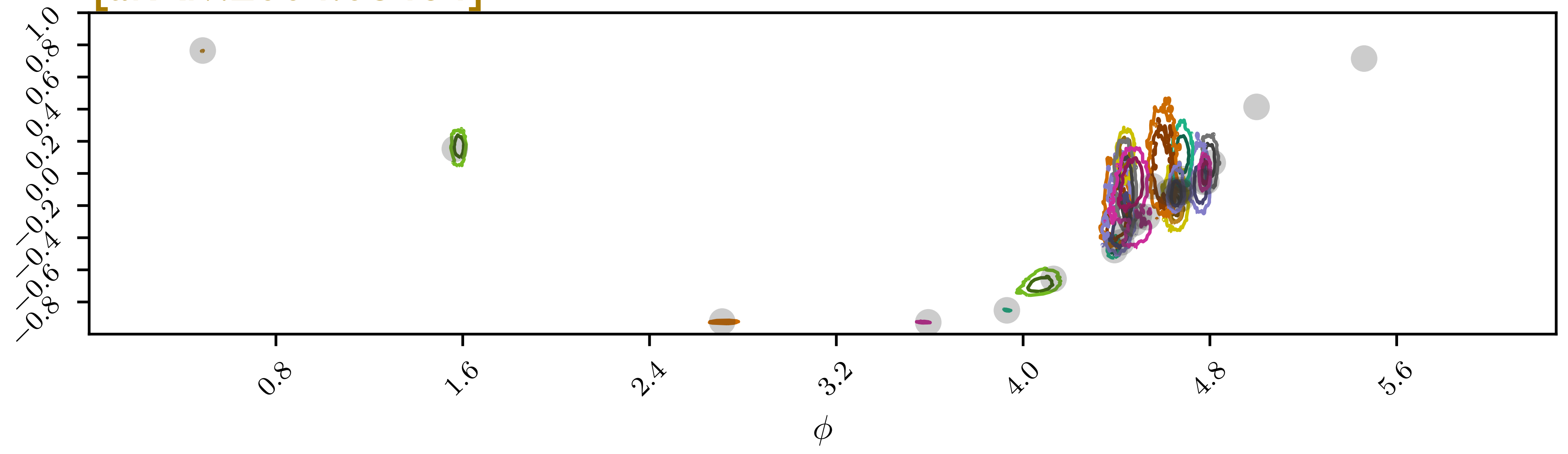
LISA: challenges for GBs



[arXiv:2004.08464]

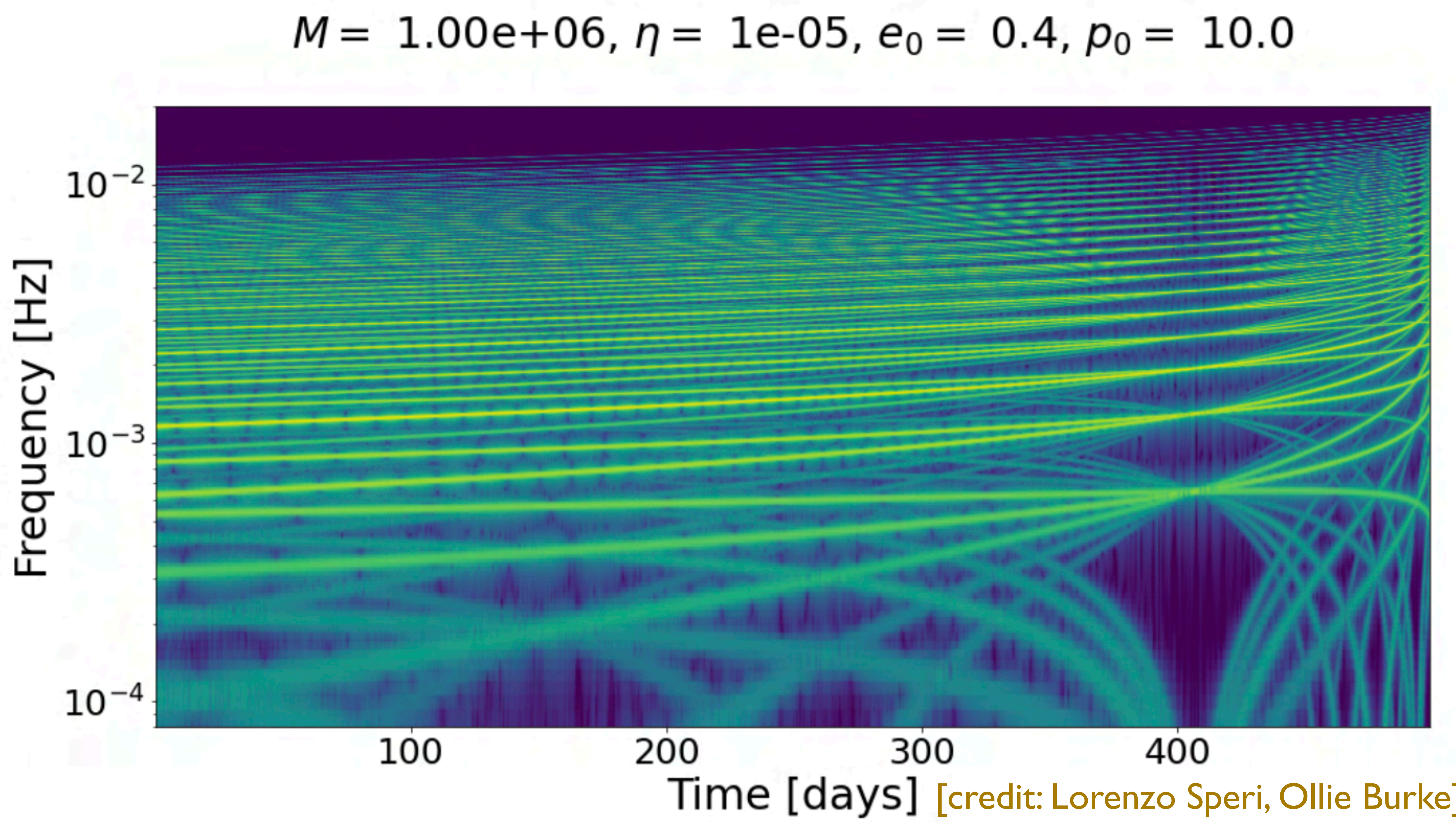


[arXiv:2004.08464]

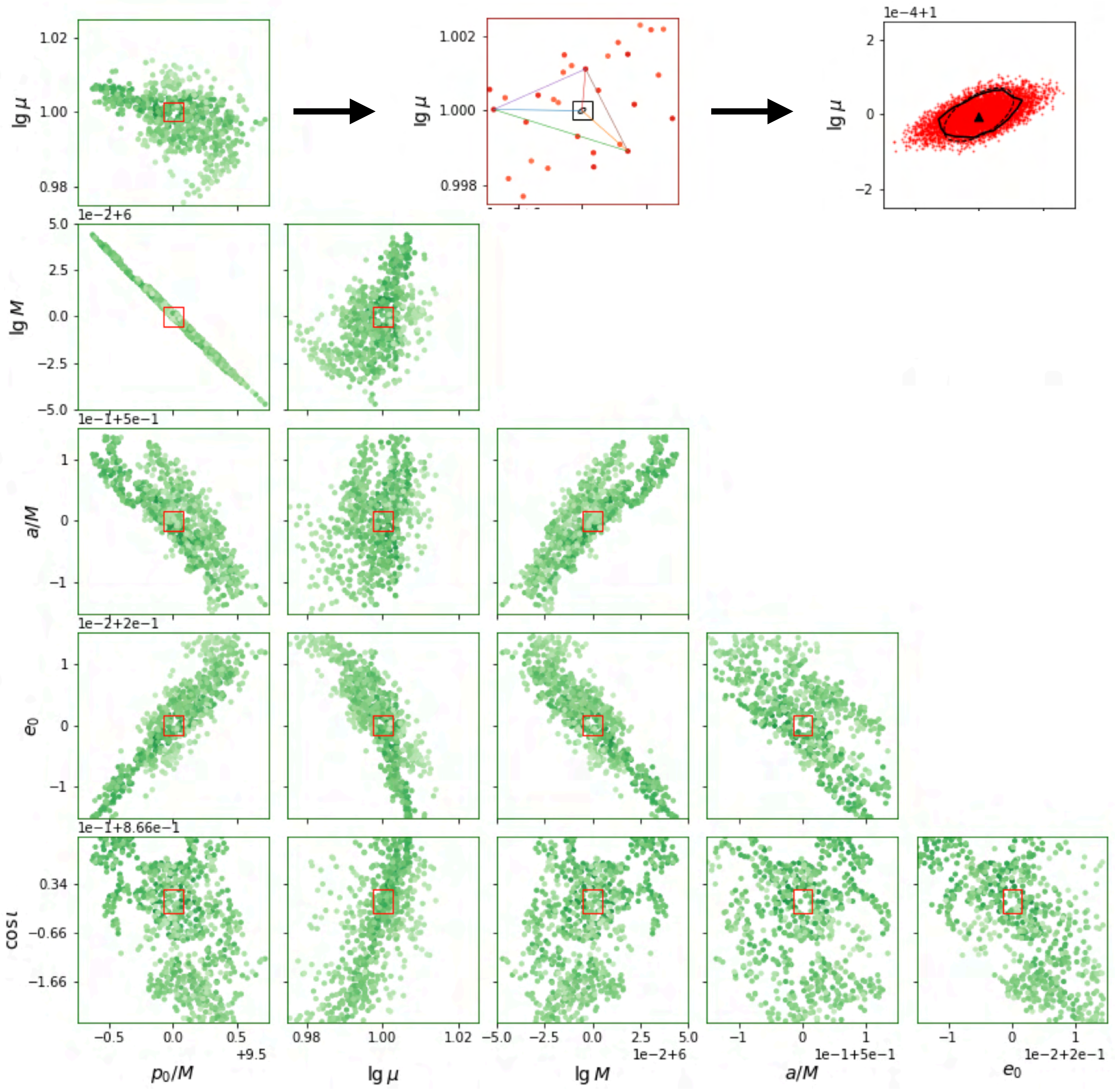


- Number of signals: millions in total, ~10000 resolvable !
- Important source confusion in the middle of the frequency range
- Techniques: transdimensional MCMC (Reversible Jump MCMC)

LISA: challenges for EMRIs



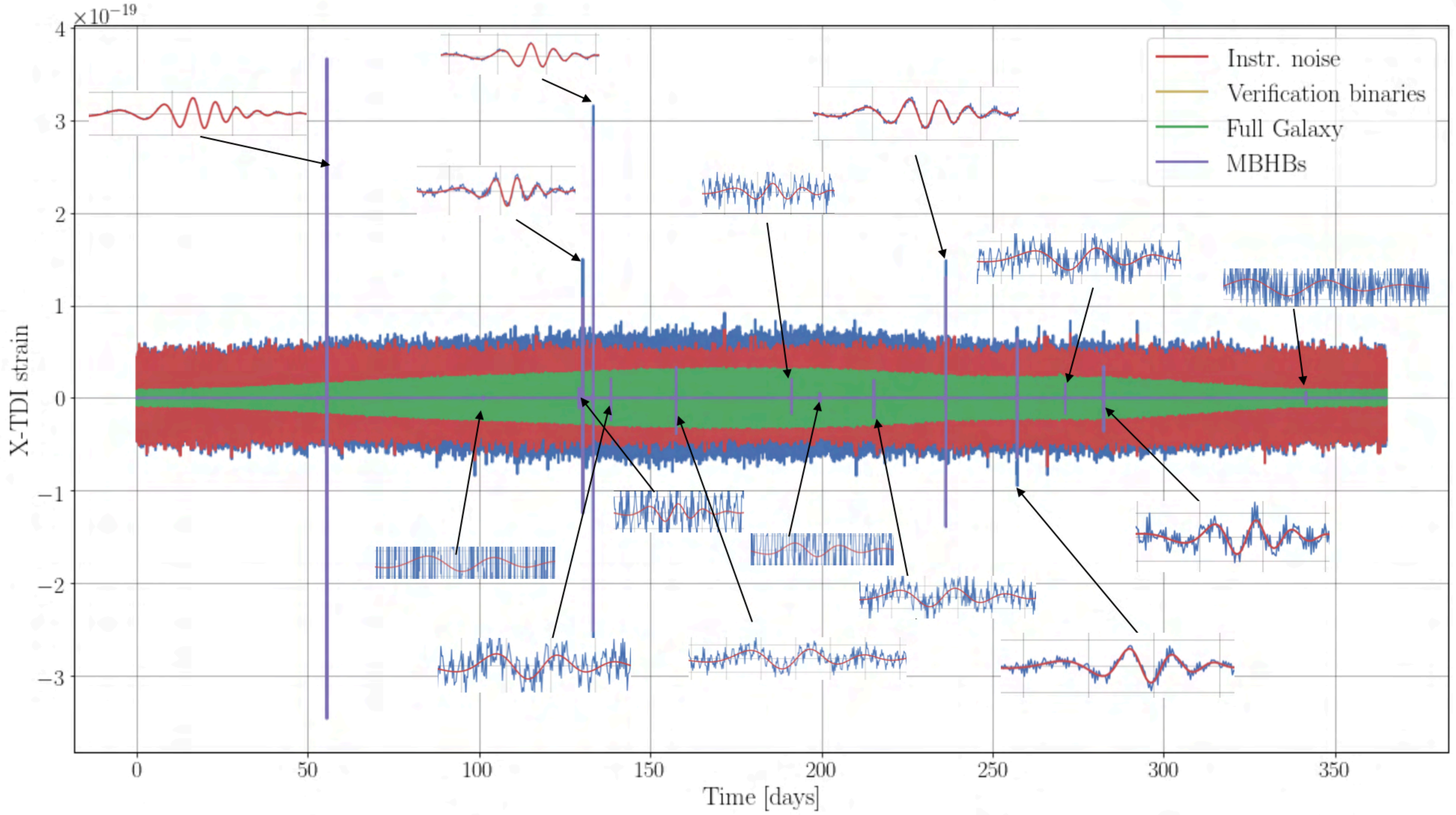
[credit: Lorenzo Speri, Ollie Burke]

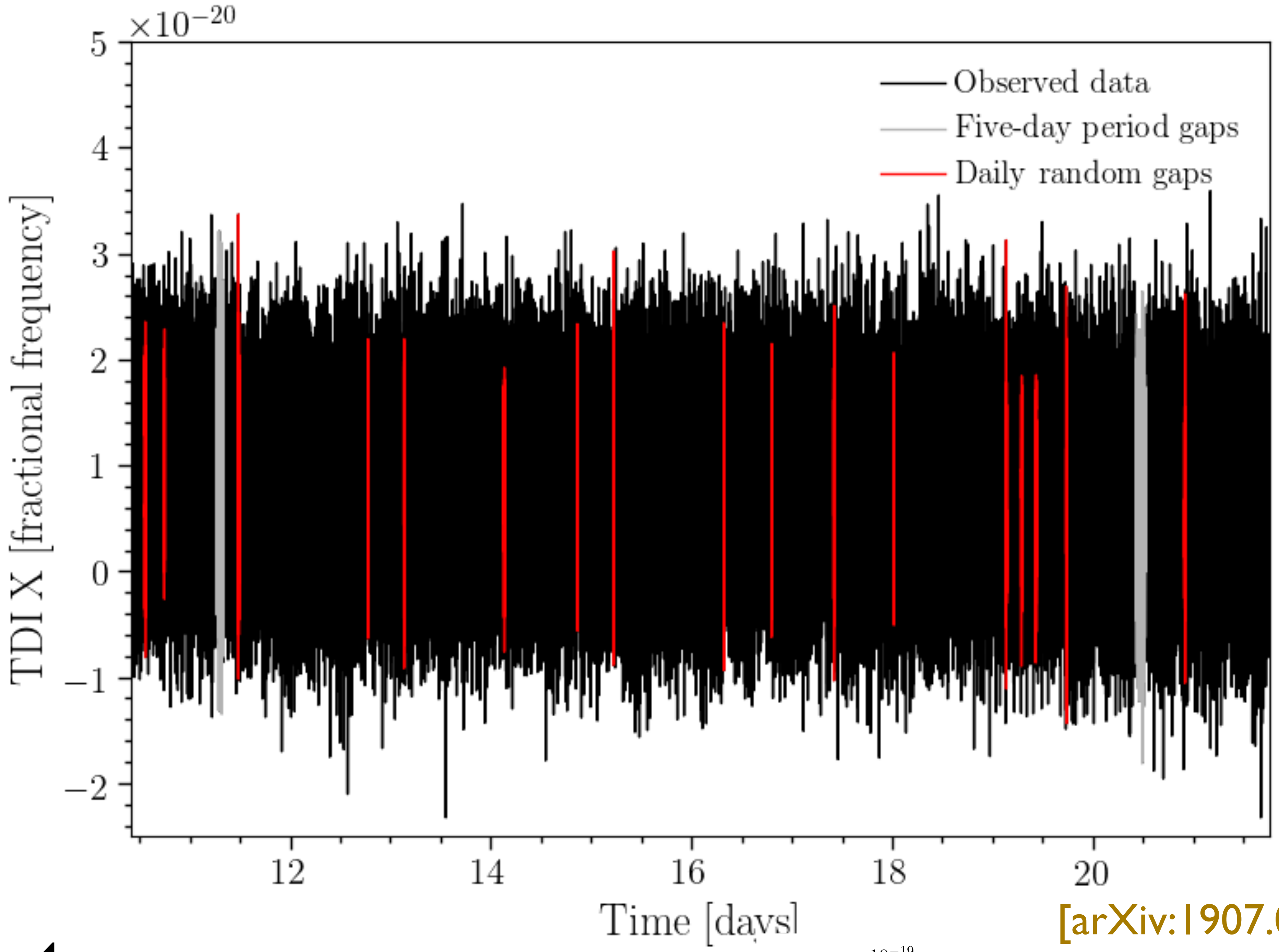


[arXiv:2109.14254]

- Complex signals, modelled in perturbative GR (frontier: 2nd order self-force)
- Long-lived signals
- Very rich harmonic structure
- Strong multimodality in parameter space

LISA: non-stationarity and gaps





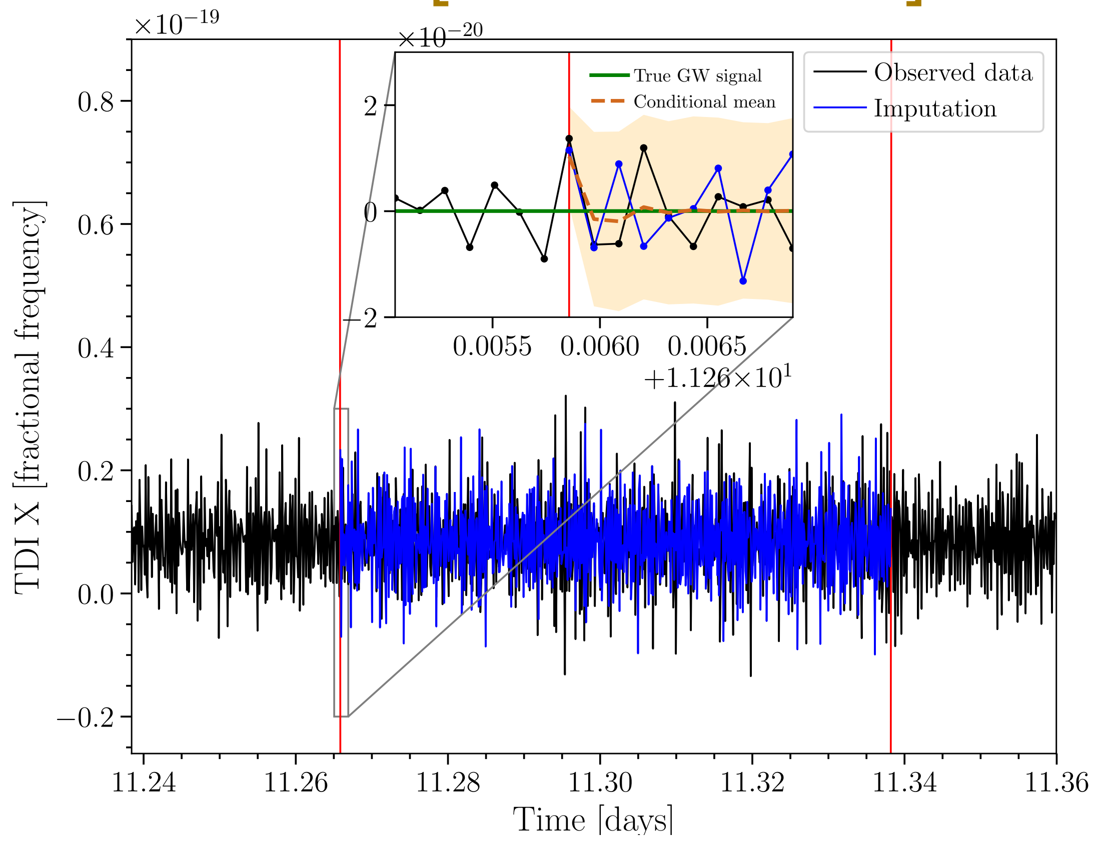
[arXiv:1907.04747]

Non-stationarity

- Non-stationarity background from double WD in the galaxy
- Instrumental non-stationarity over long times
- Glitches (as seen in LISA Pathfinder)

Data gaps

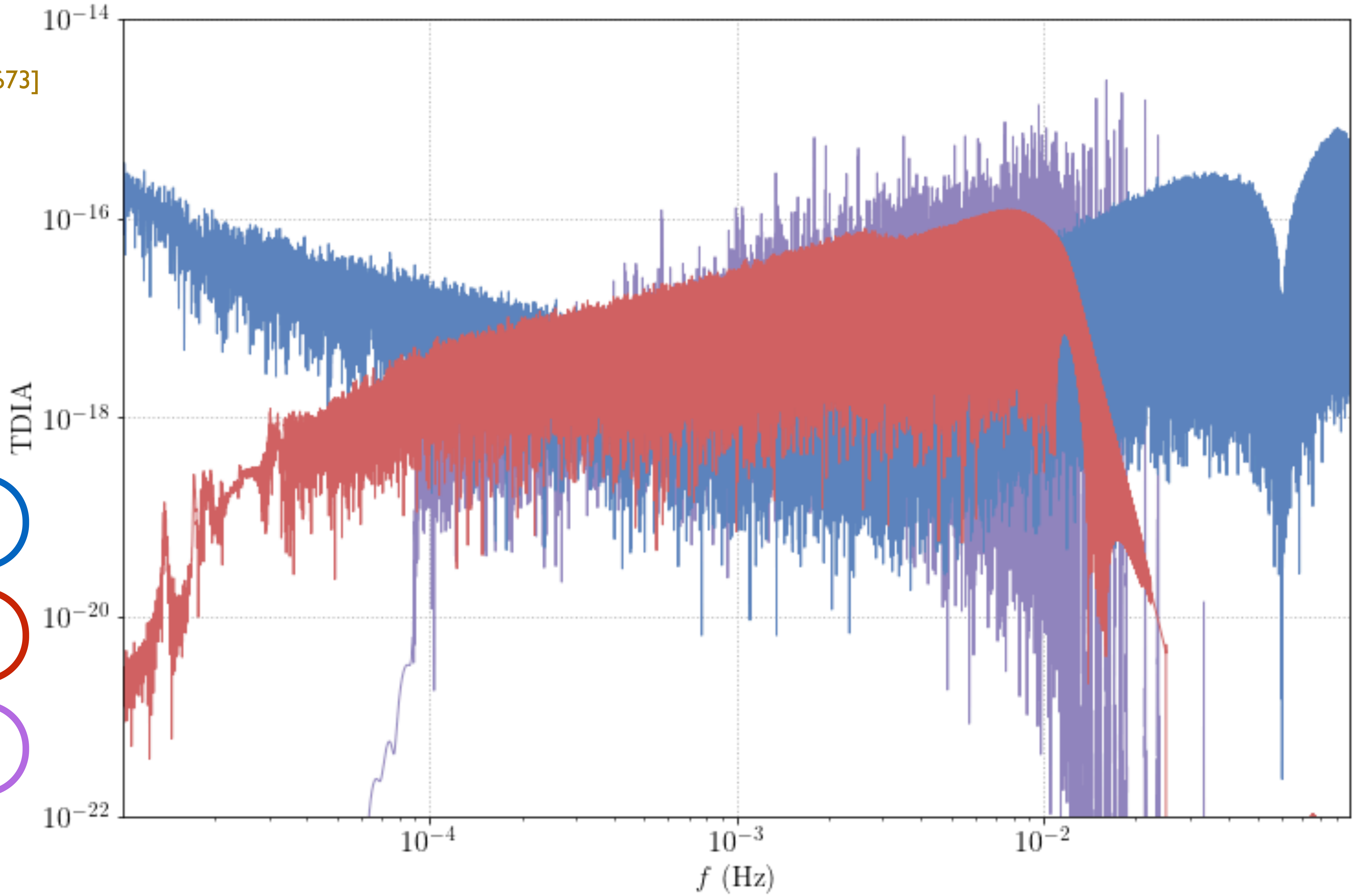
- Both scheduled and unscheduled
- Mask/taper data ?
- Inpainting methods ?



LDC Sangria: first steps of a global fit

[APC/L2IT, in prep.]

see also [arXiv:2301.03673]



Instr. noise

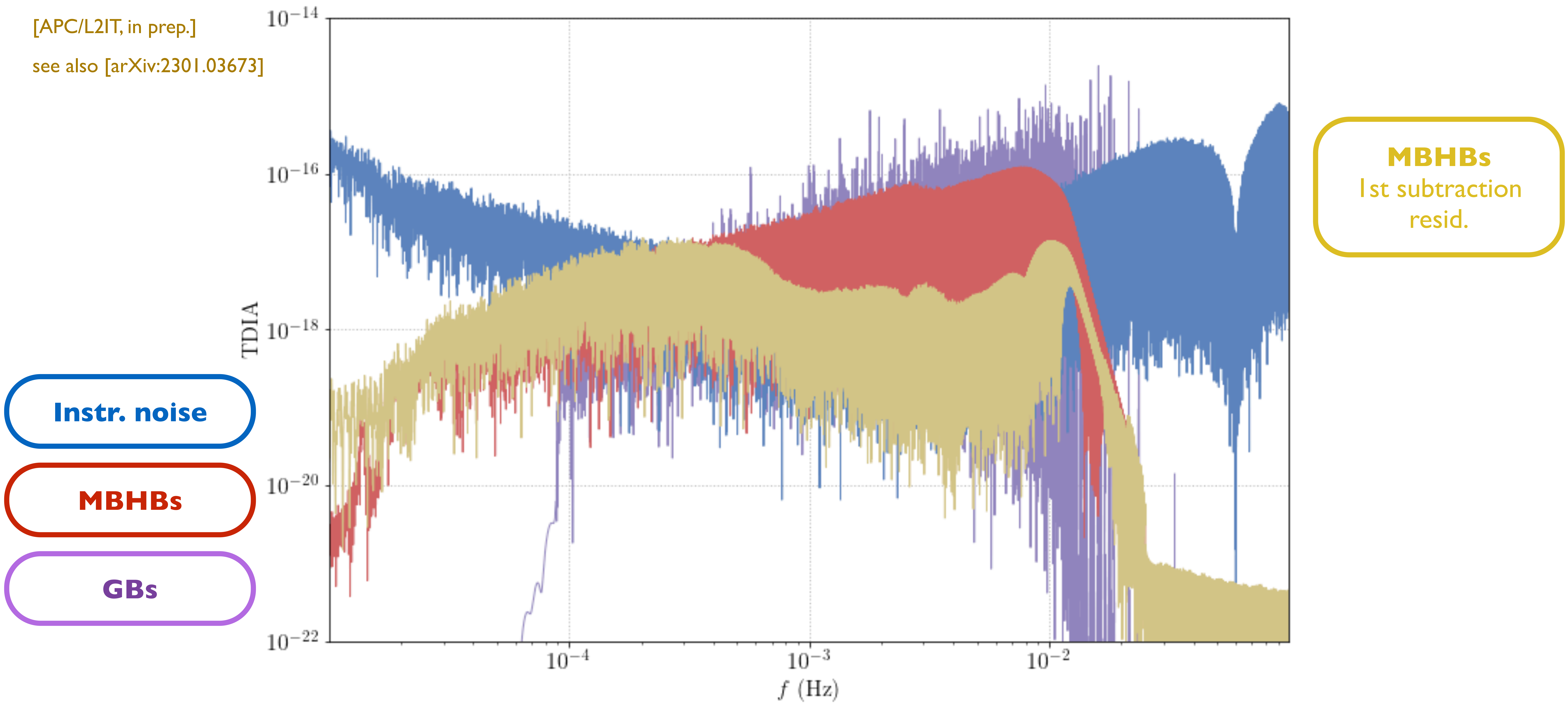
MBHBs

GBs

LDC Sangria: first steps of a global fit

[APC/L2IT, in prep.]

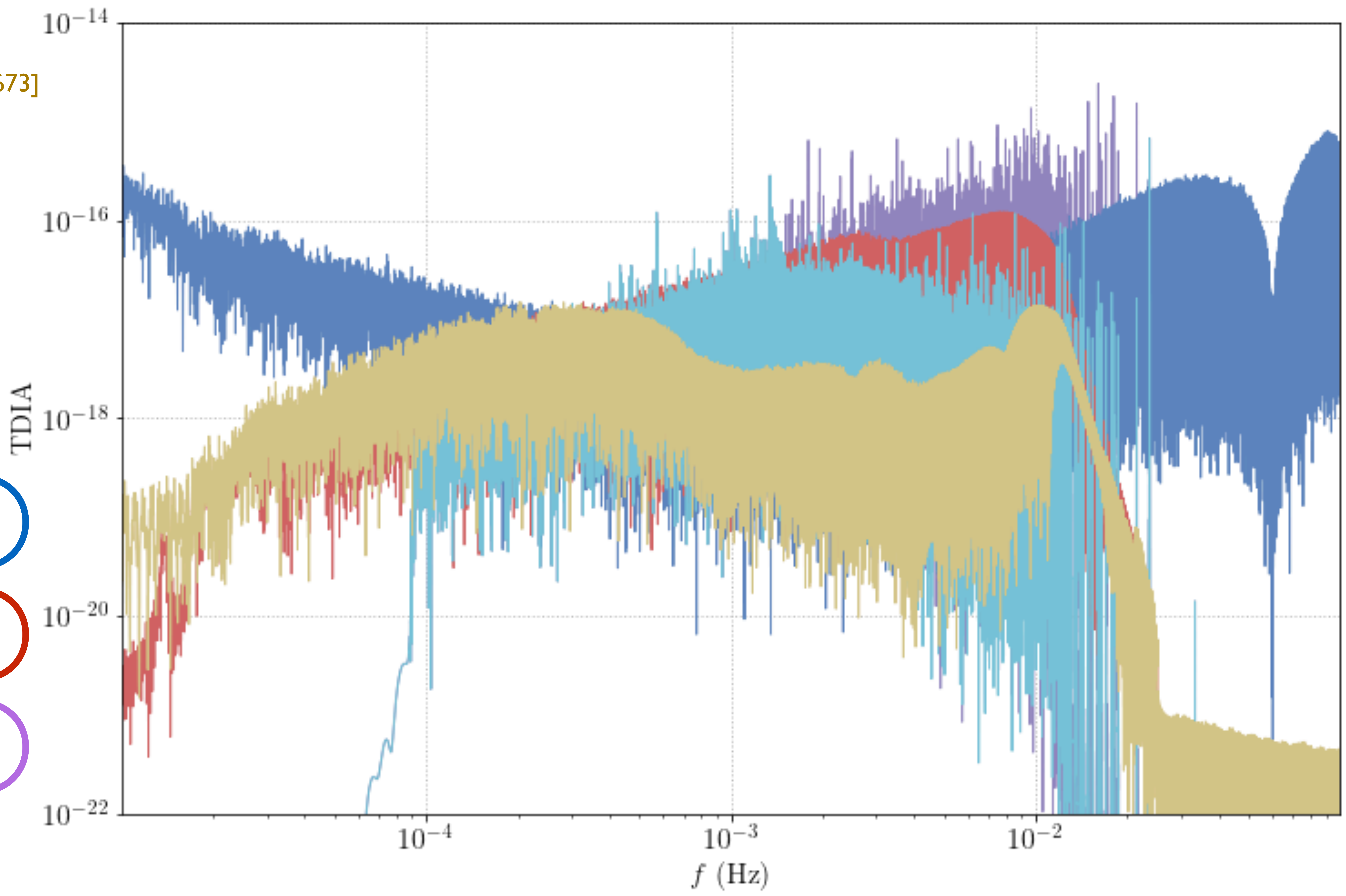
see also [arXiv:2301.03673]



LDC Sangria: first steps of a global fit

[APC/L2IT, in prep.]

see also [arXiv:2301.03673]



Instr. noise

MBHBs

GBs

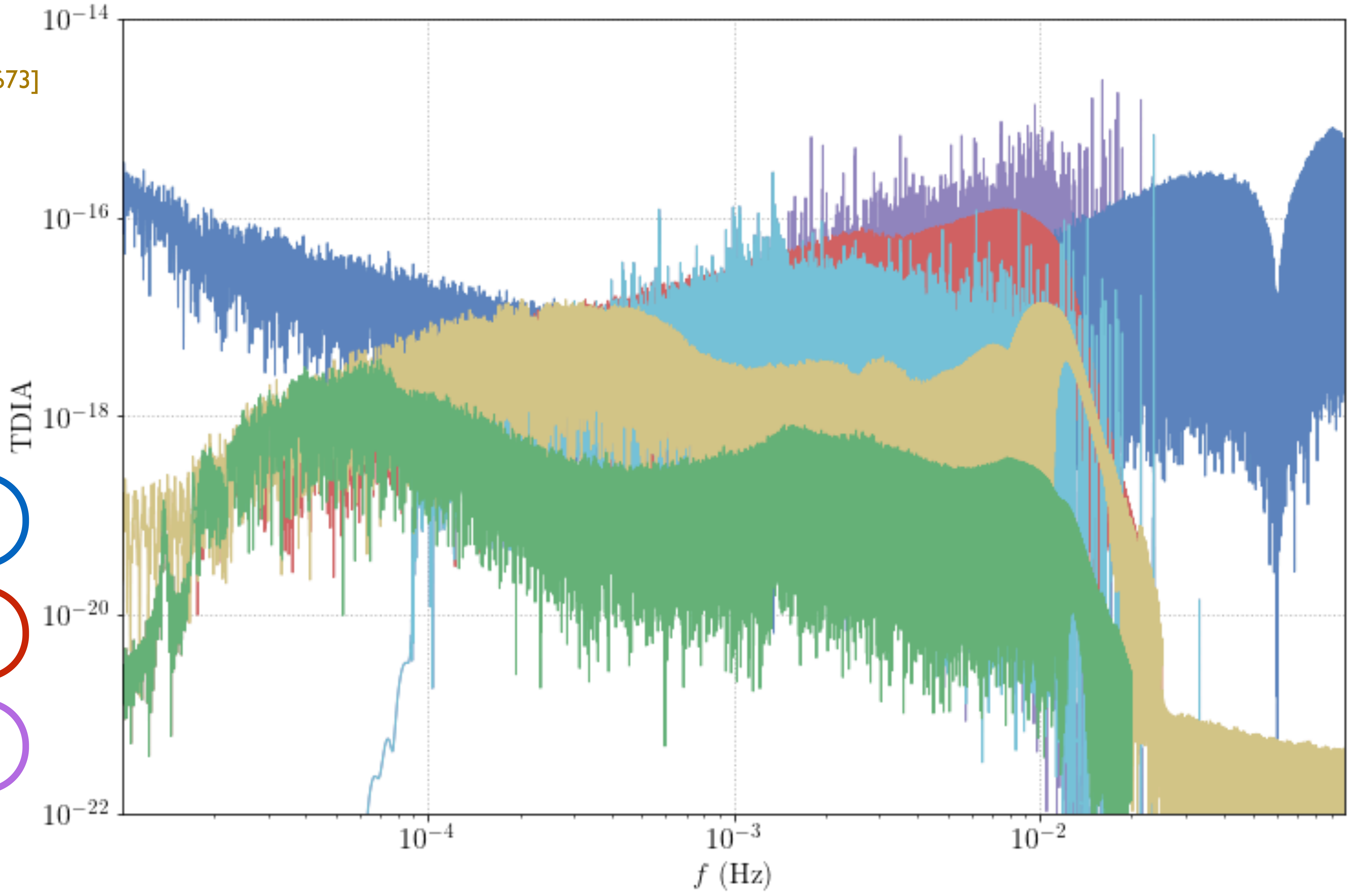
MBHBs
1st subtraction resid.

GBs
1st PE resid.

LDC Sangria: first steps of a global fit

[APC/L2IT, in prep.]

see also [arXiv:2301.03673]



Instr. noise

MBHBs

GBs

MBHBs
1st subtraction resid.

GBs
1st PE resid.

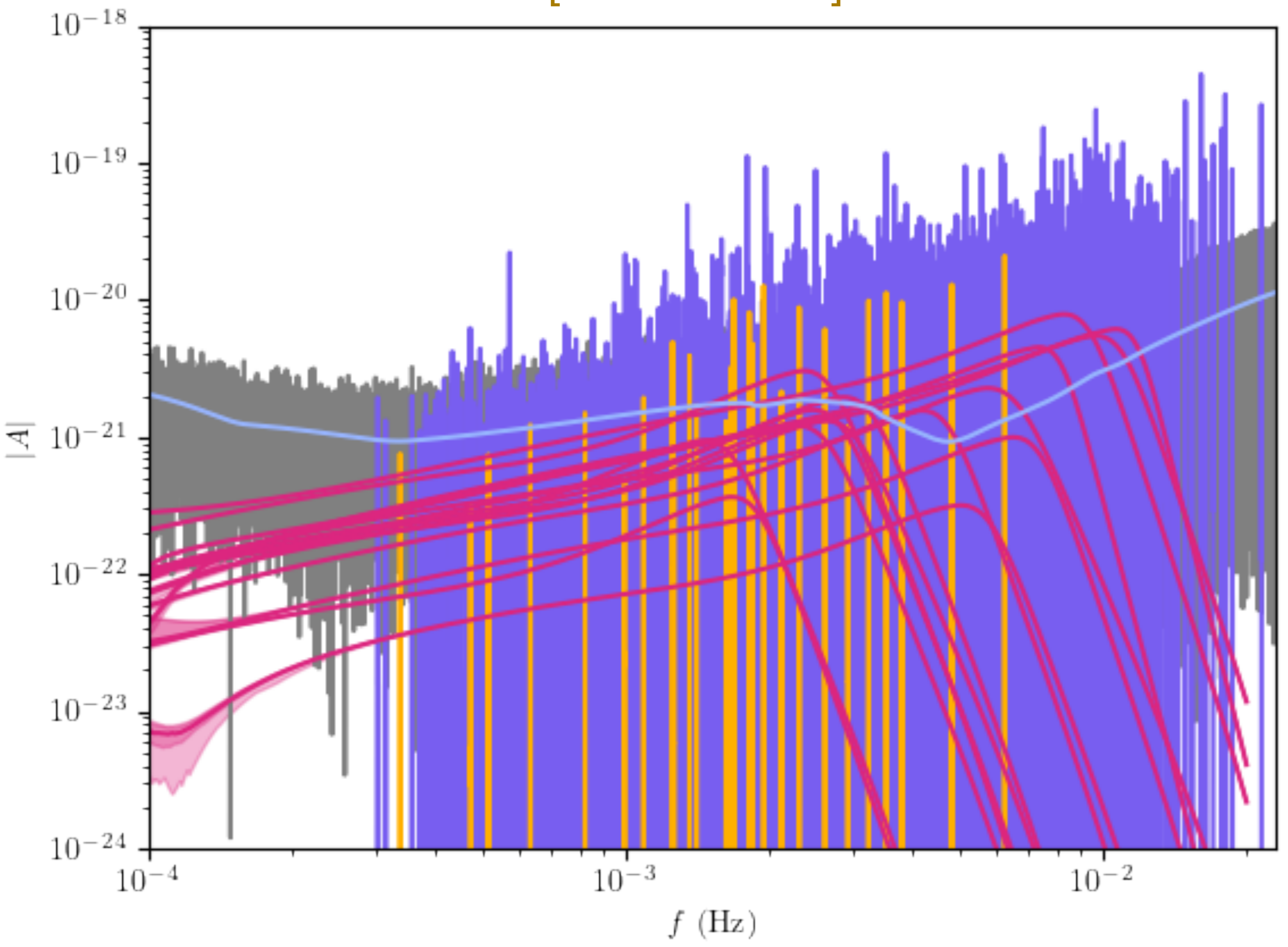
MBHBs
1st PE resid.

+ 2nd PE GBs

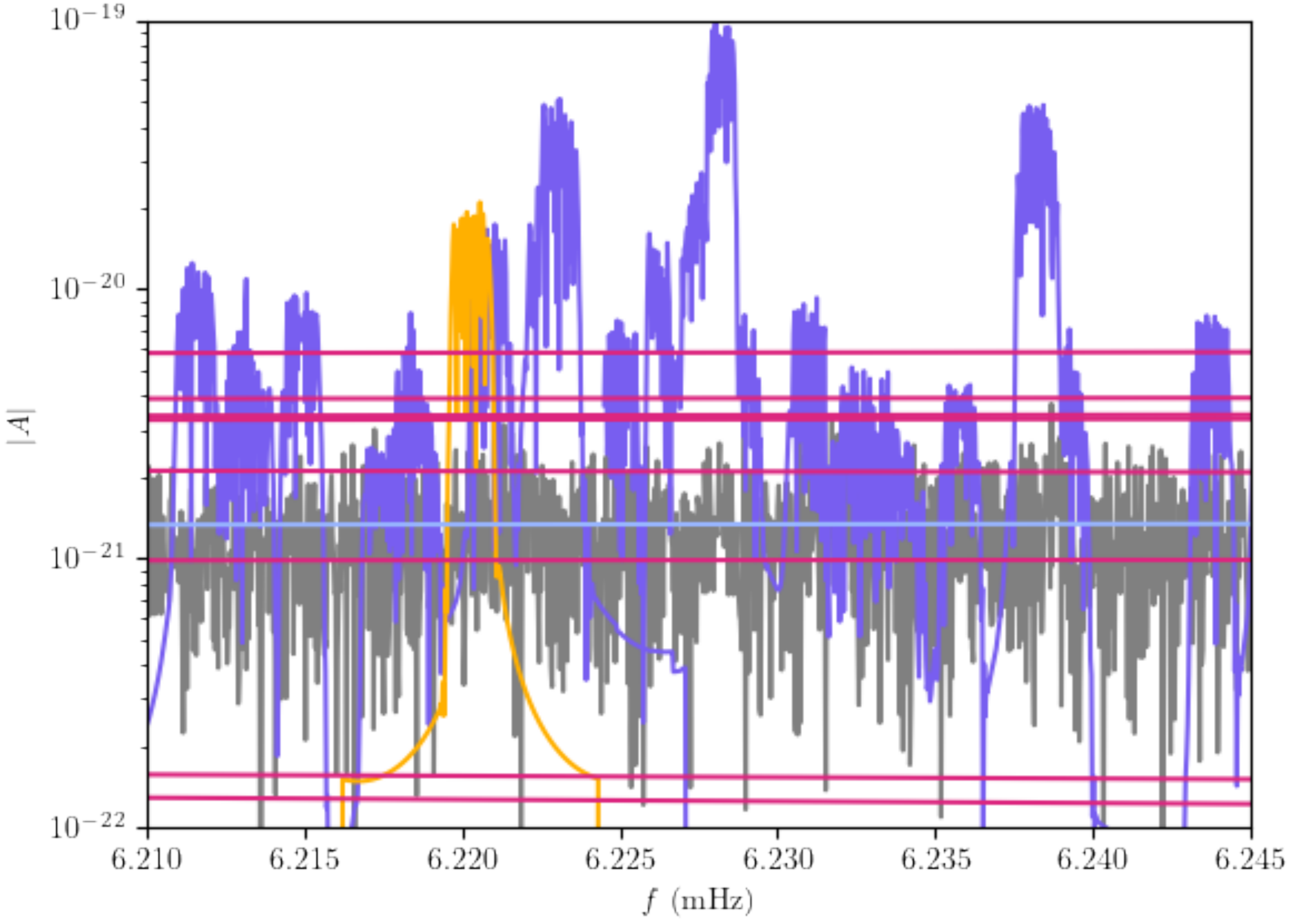
+ 2nd PE MBHBs

LDC Sangria: preview of results

[arXiv:2301.03673]



All signals recovered FD

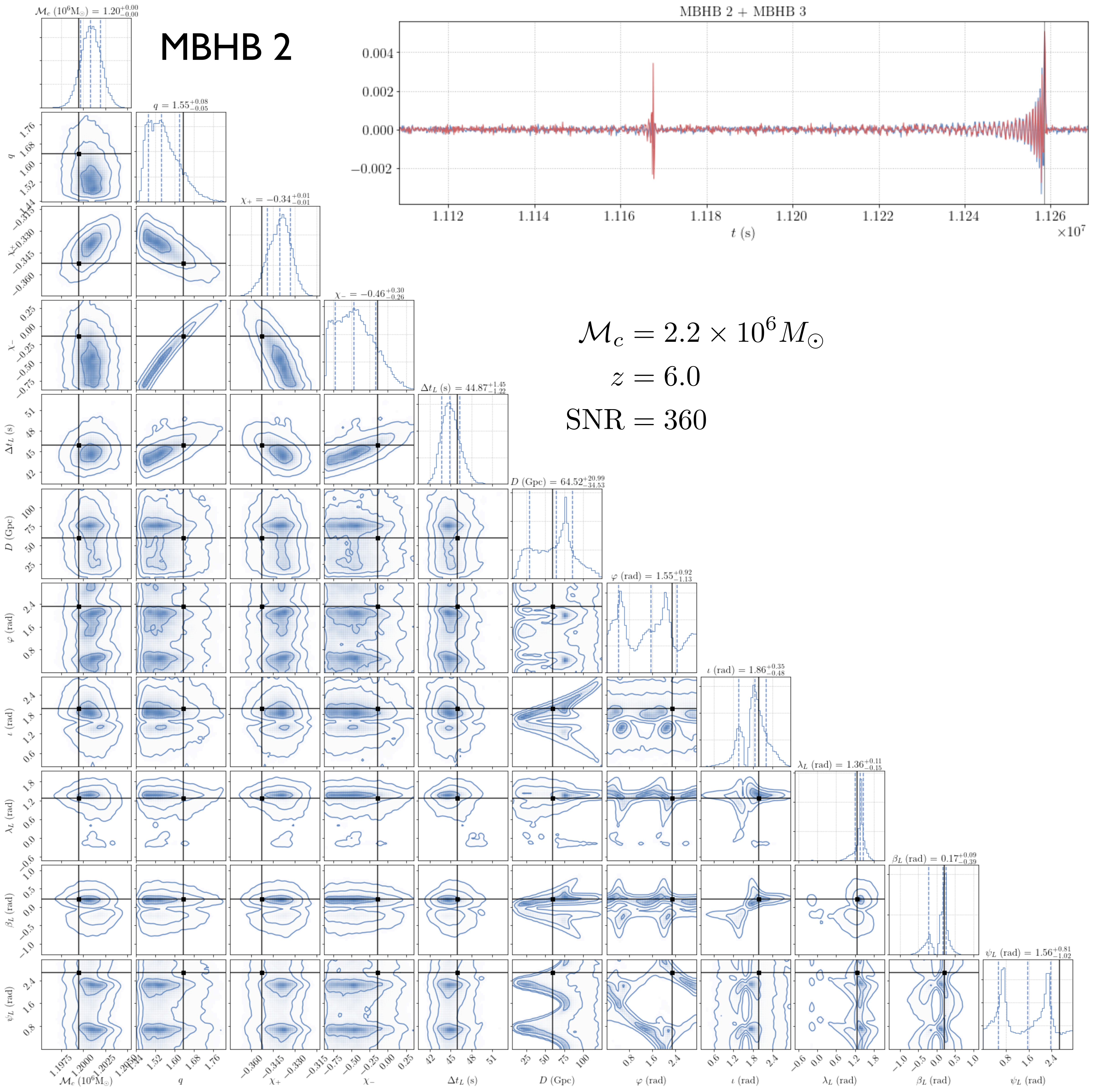


Superposed GB identification

LDC Sangria: preview of results

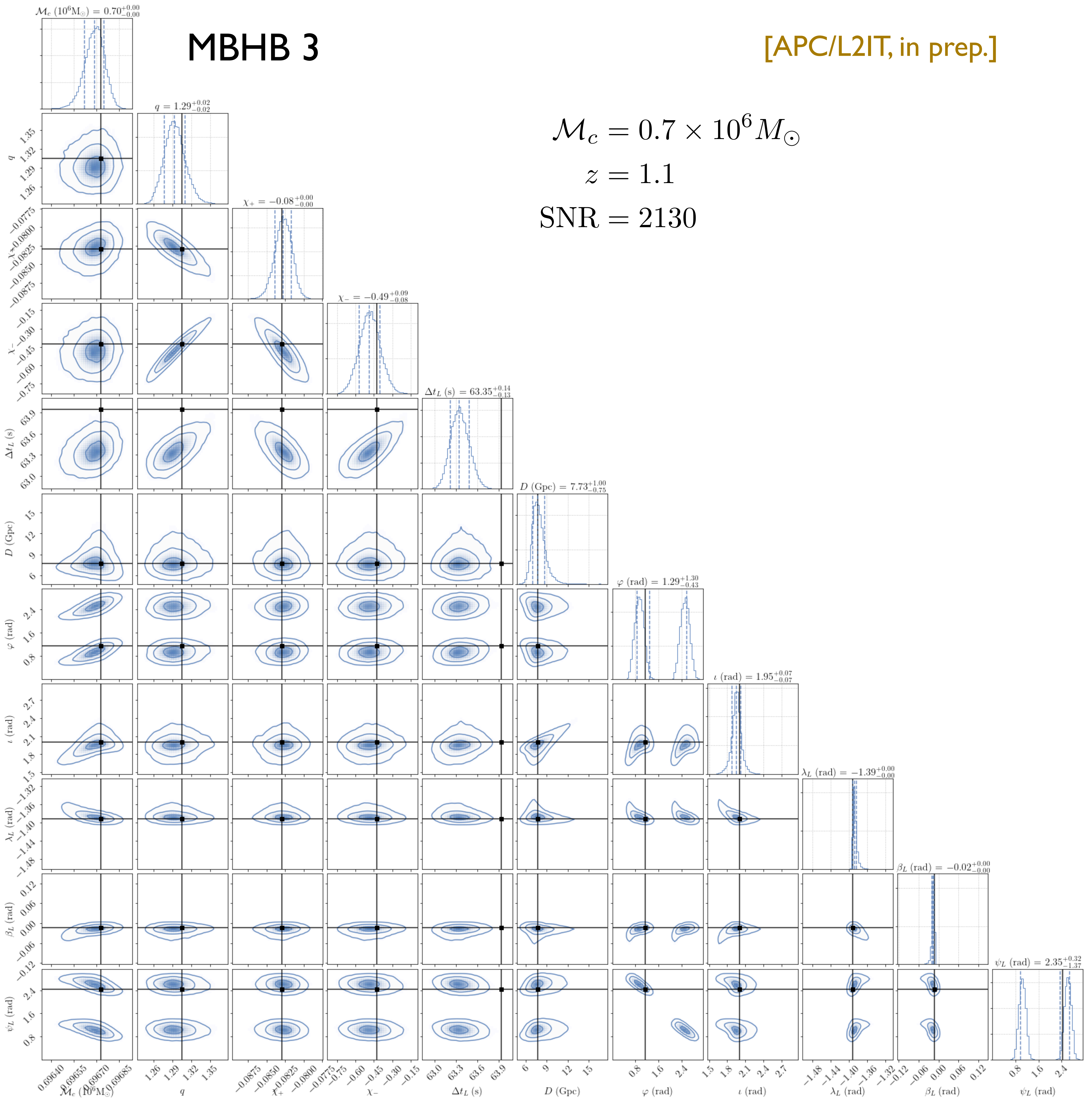
[APC/L2IT, in prep.]

MBHB 2



$$\mathcal{M}_c = 2.2 \times 10^6 M_\odot$$
$$z = 6.0$$
$$\mathrm{SNR} = 360$$

MBHB 3

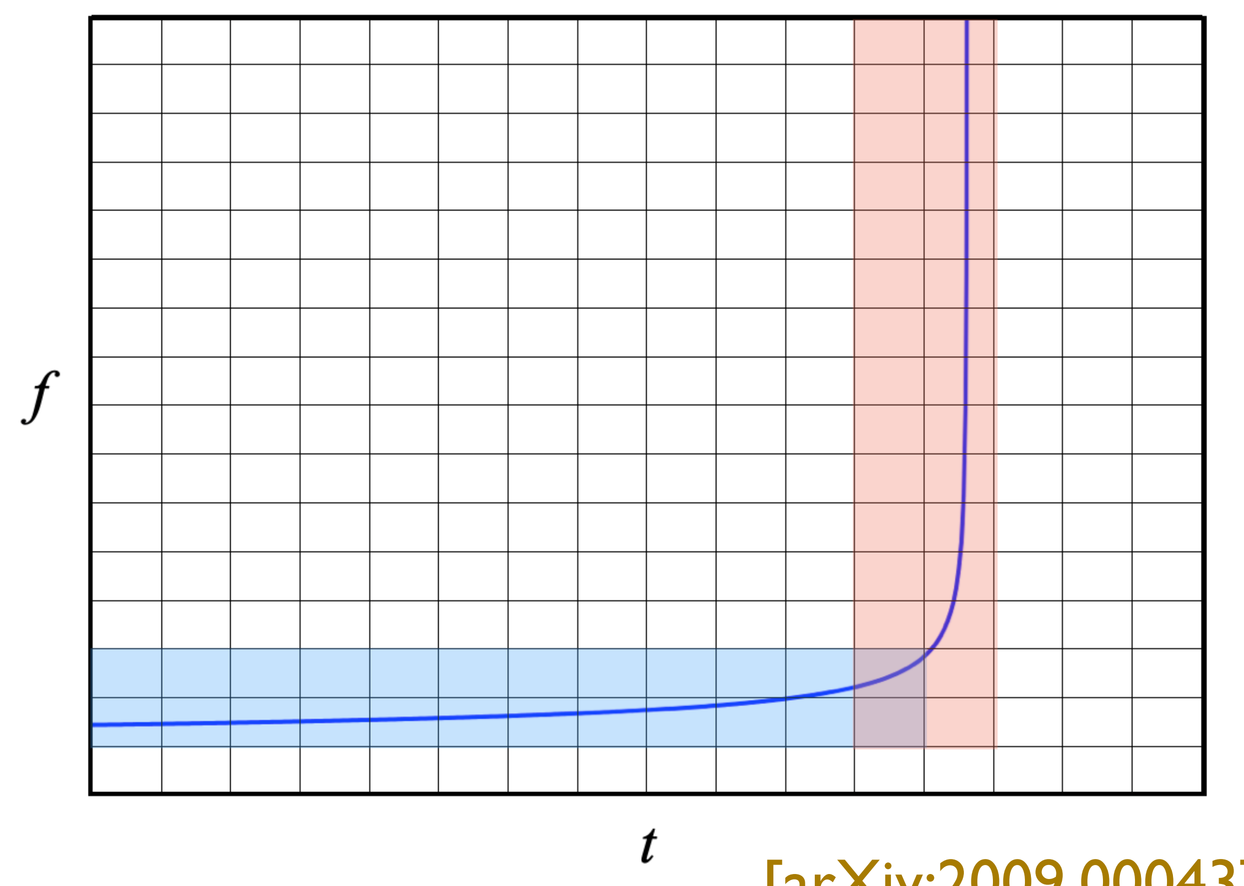


$$\mathcal{M}_c = 0.7 \times 10^6 M_\odot$$
$$z = 1.1$$
$$\mathrm{SNR} = 2130$$

The future of GWDA

Wavelet domain

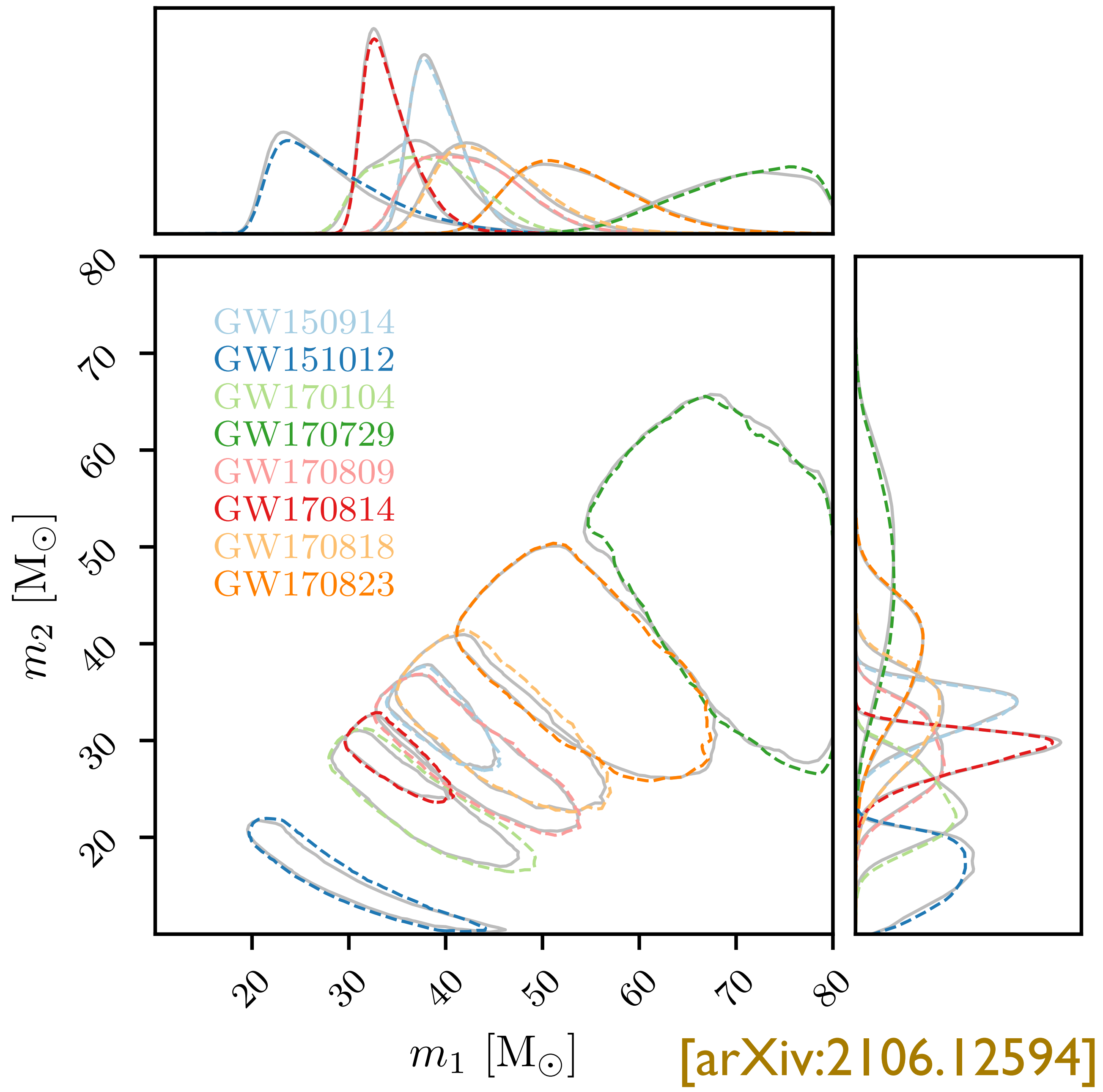
- more compact representation for chirps
- fast chirplet transform instead of DFT
- natural framework for non-stationarity



[arXiv:2009.00043]

Machine learning

- applications to glitch identification
- applications in waveform modelling
- simulation-based inference for PE



[arXiv:2106.12594]

GPUs

- computation paradigm of the future
- see also new languages with autograd, Jax