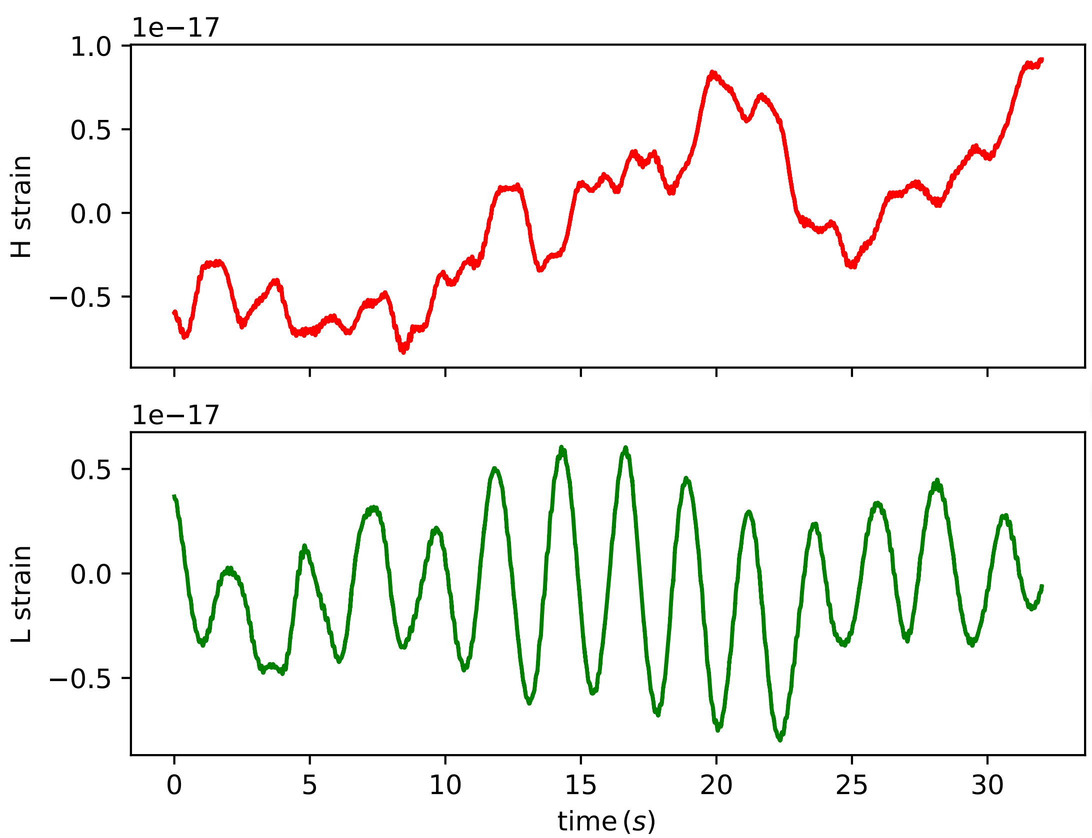
Gravitational waves data analysis

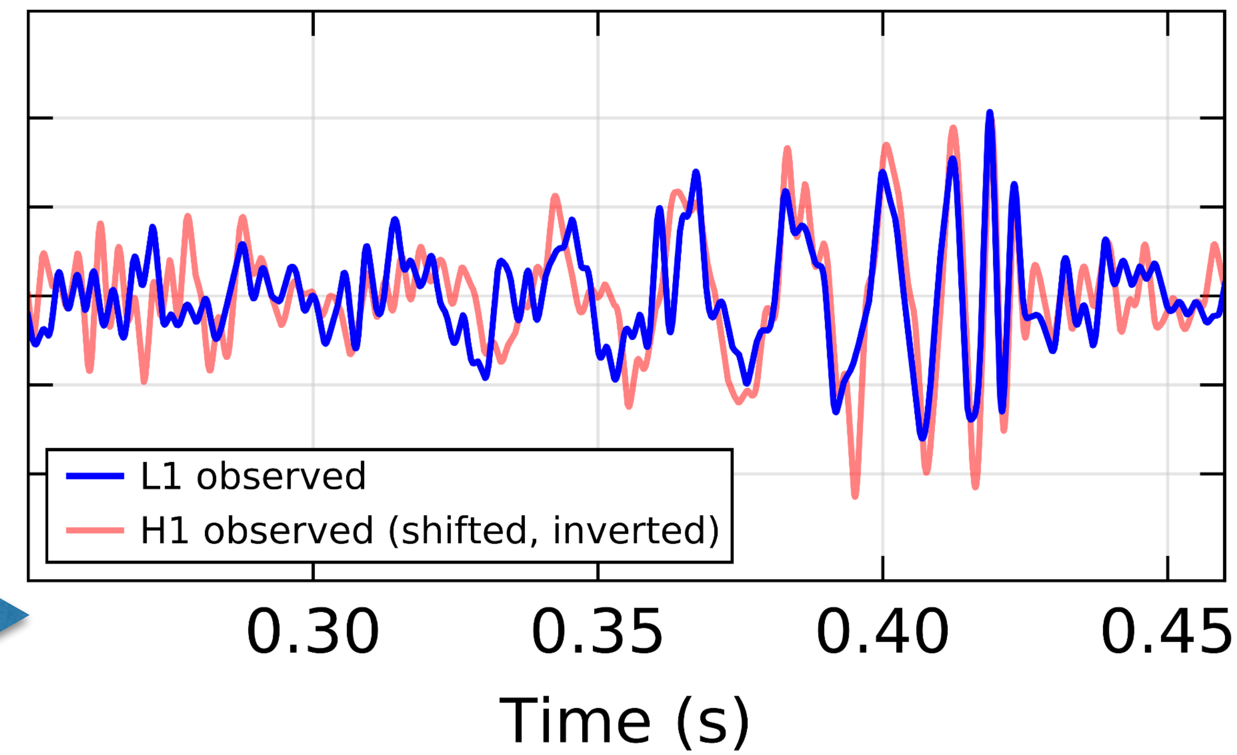
Walter Del Pozzo

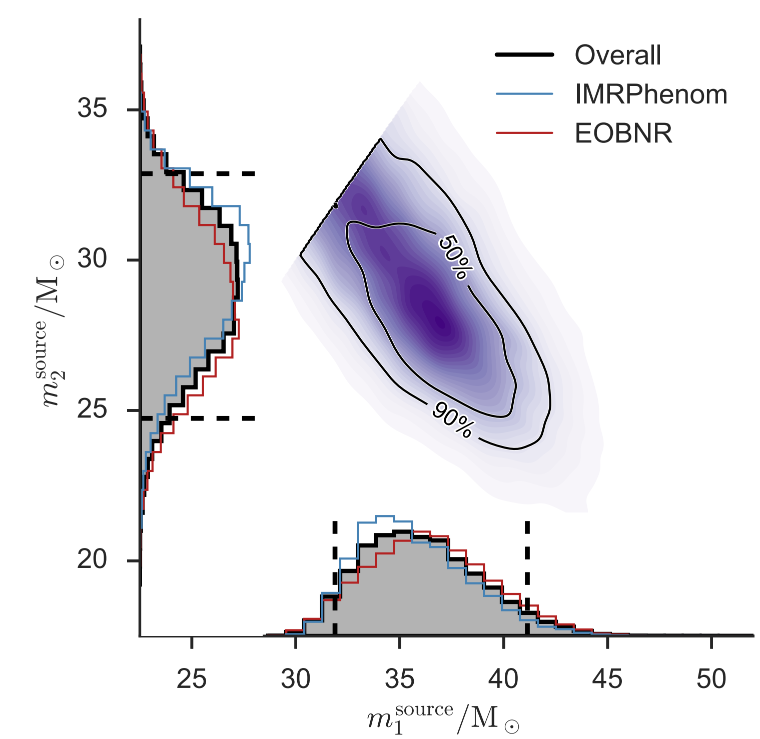
Outline

- Part 1: fundaments of probability theory
- Part 2: stochastic processes
- Part 3: (modelled) detection of gravitational wave signals
- Part 4: inference of gravitational wave physics
- Part 5: overview of (selected) tests of general relativity

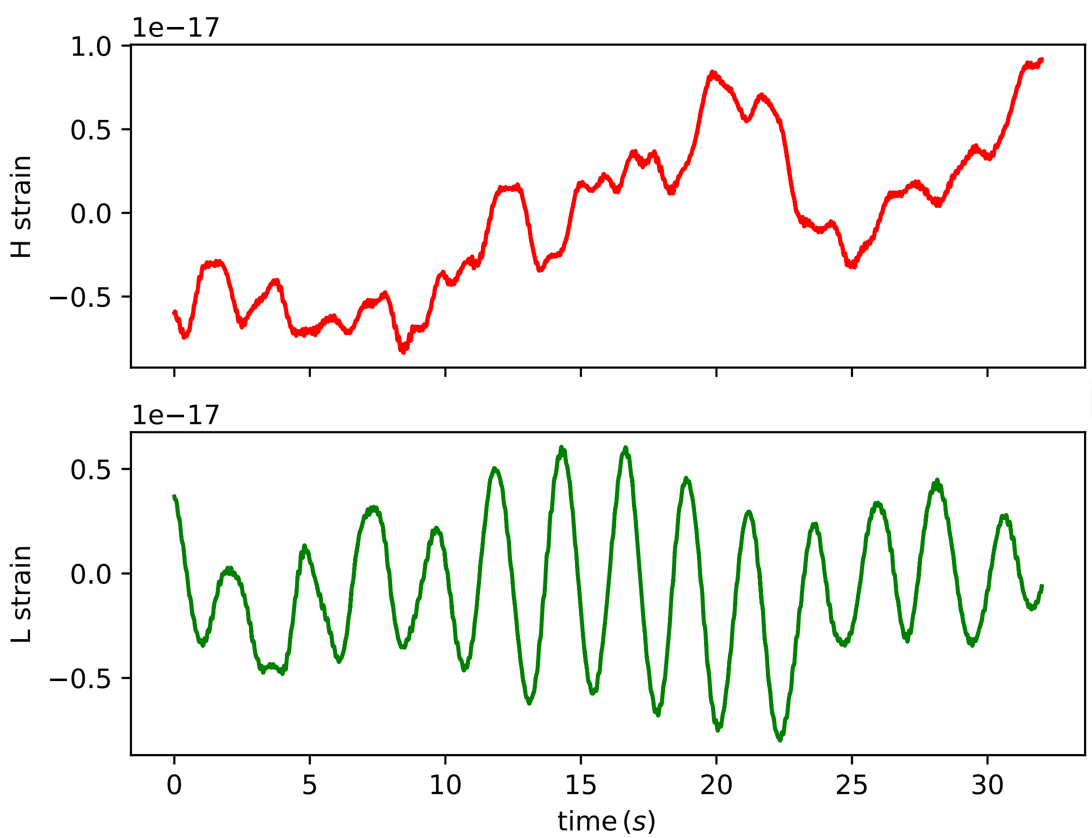
What is “data analysis”?

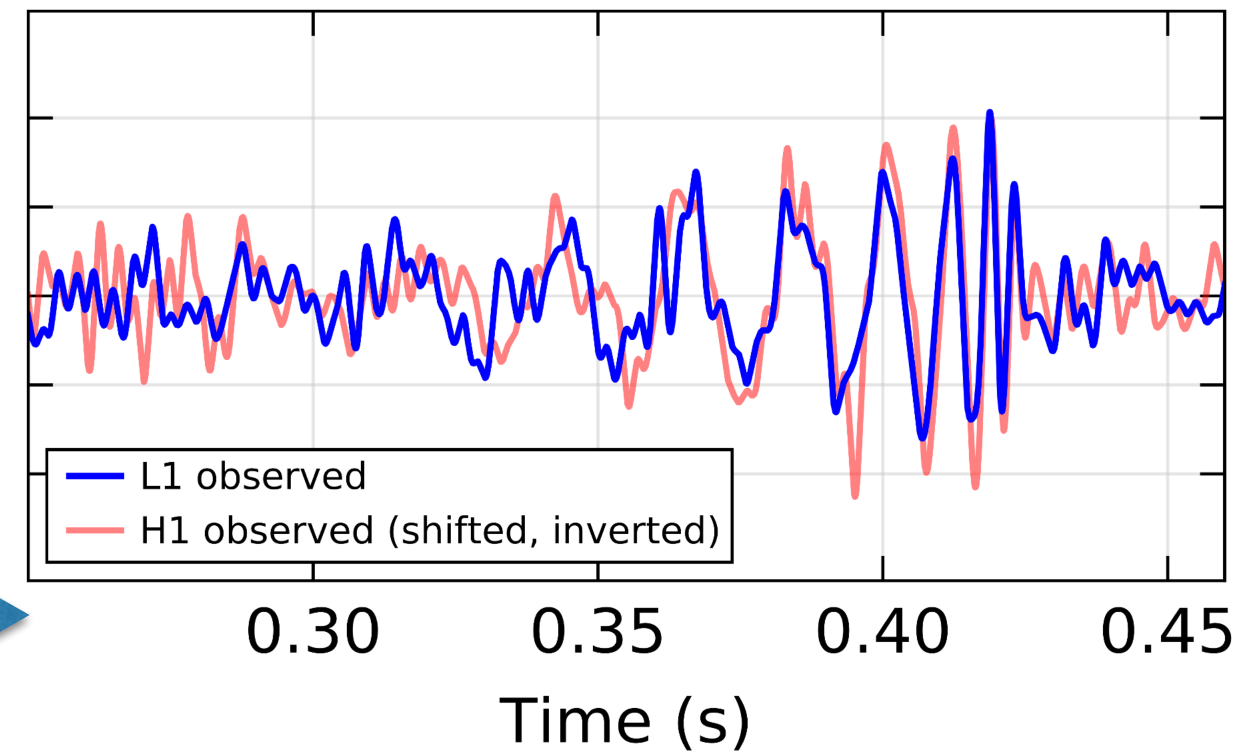


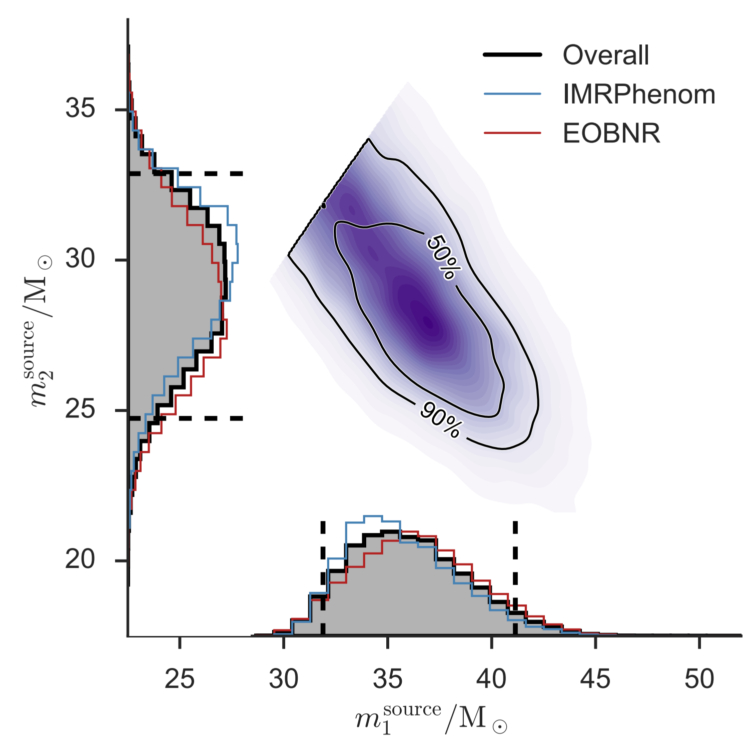




What is “data analysis”?



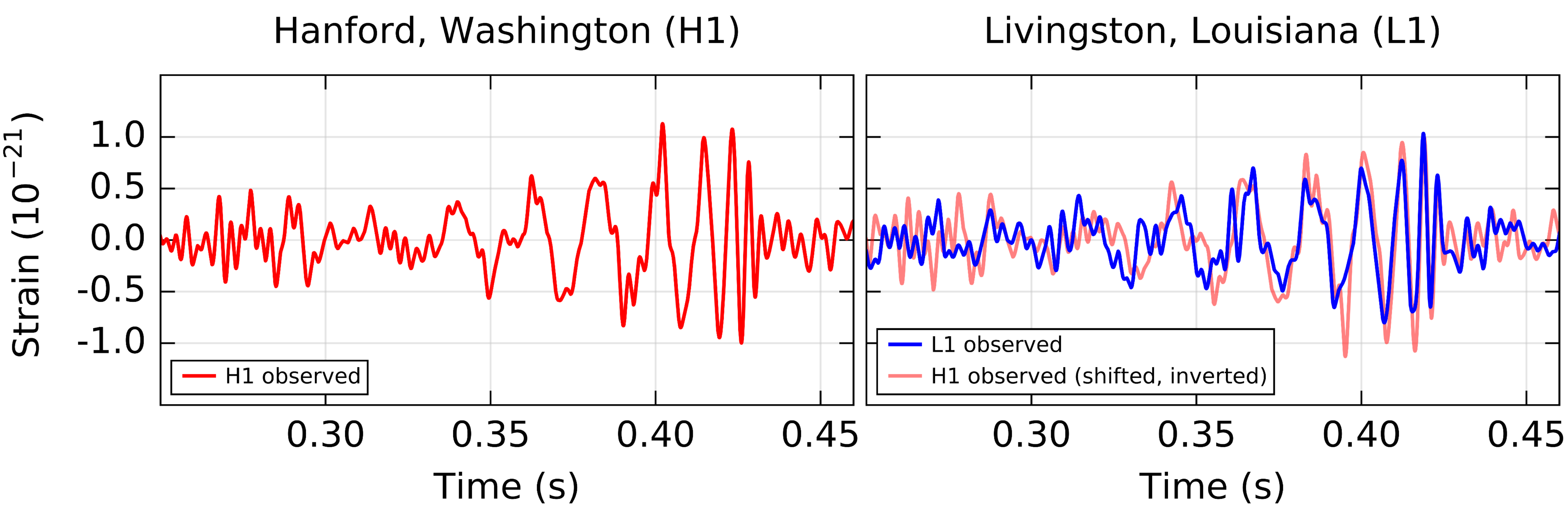




$$p(H|DI) = p(H|I)\frac{p(D|HI)}{p(D|I)}$$

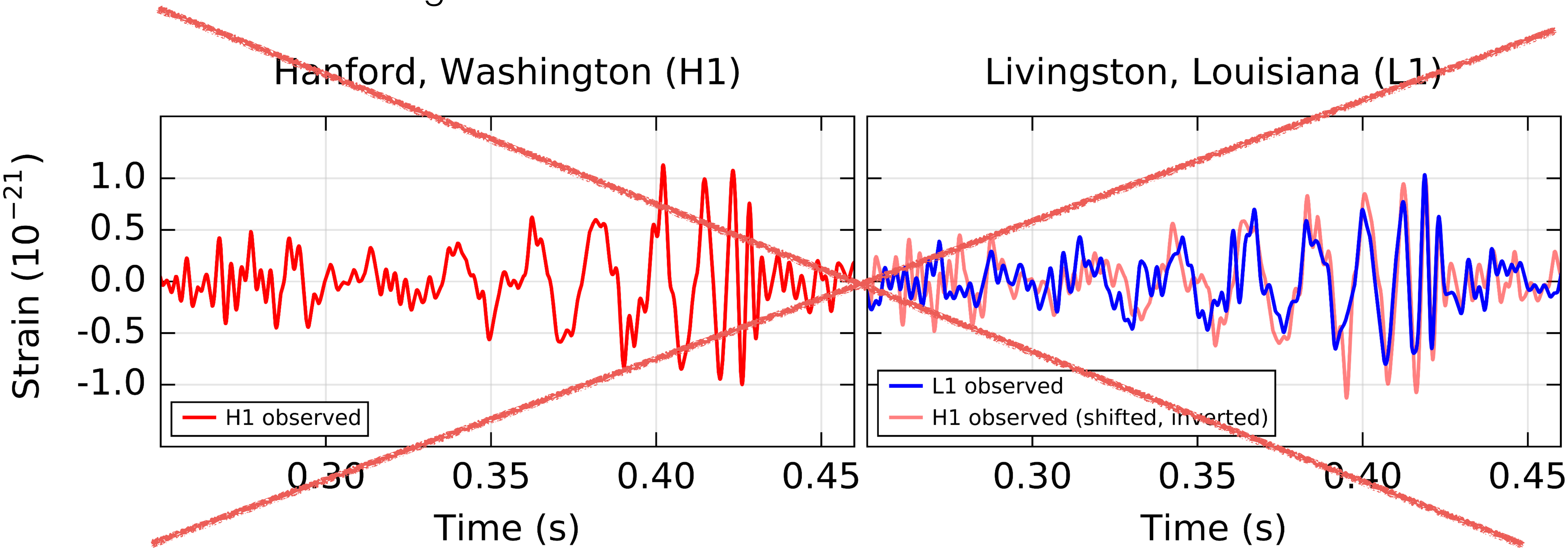
Why Bayes' theorem?

- Gravitational wave events are rare
- Noise dominated detectors
- Need to know what we are looking for VERY well to detect it/measure its properties
  - Matched filtering

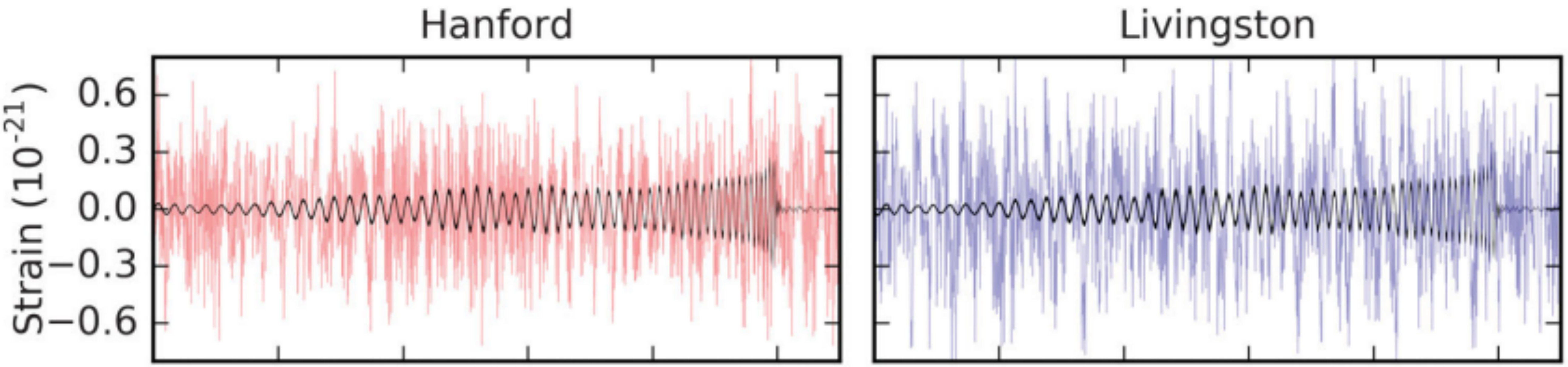


Why Bayes' theorem?

- Gravitational wave events are rare
- Noise dominated detectors
- Need to know what we are looking for VERY well to detect it/measure its properties
  - Matched filtering



Why Bayes' theorem?



- Need for accurate predictions to dig into the noise floor
  - Faithful models for the coalescence of compact binaries
    - effective-one-body family
    - numerical relativity
    - phenomenological models
    - surrogate models

Part 1

Fundaments of probability theory

“Probability is a property of the observer”

Fundaments

- Logical propositions:
  - Statements to which one can assign a True (1) or False (0) value
  - A = “The sun is a star”
  - B = “The total mass of GW150914 was 70 Msun”

Logical operations

- Arbitrarily complex sentences constructed via basic operations
  - Negation (NOT): $\bar{A}$
  - Conjunction (AND): $A \cdot B \equiv AB$
  - Disjunction (OR): $A + B$

Inference

- Deductive inference
  - Strong syllogism:
    - major premise: if A is true then B is true
    - minor premise: A is true
    - conclusion: B is true

$$(A = AB) \equiv (A \implies B)$$

- Inductive inference
  - Weak syllogism:
    - major premise: if A is true then B is true
    - minor premise: B is true
    - conclusion: A is more plausible

Deductive inference

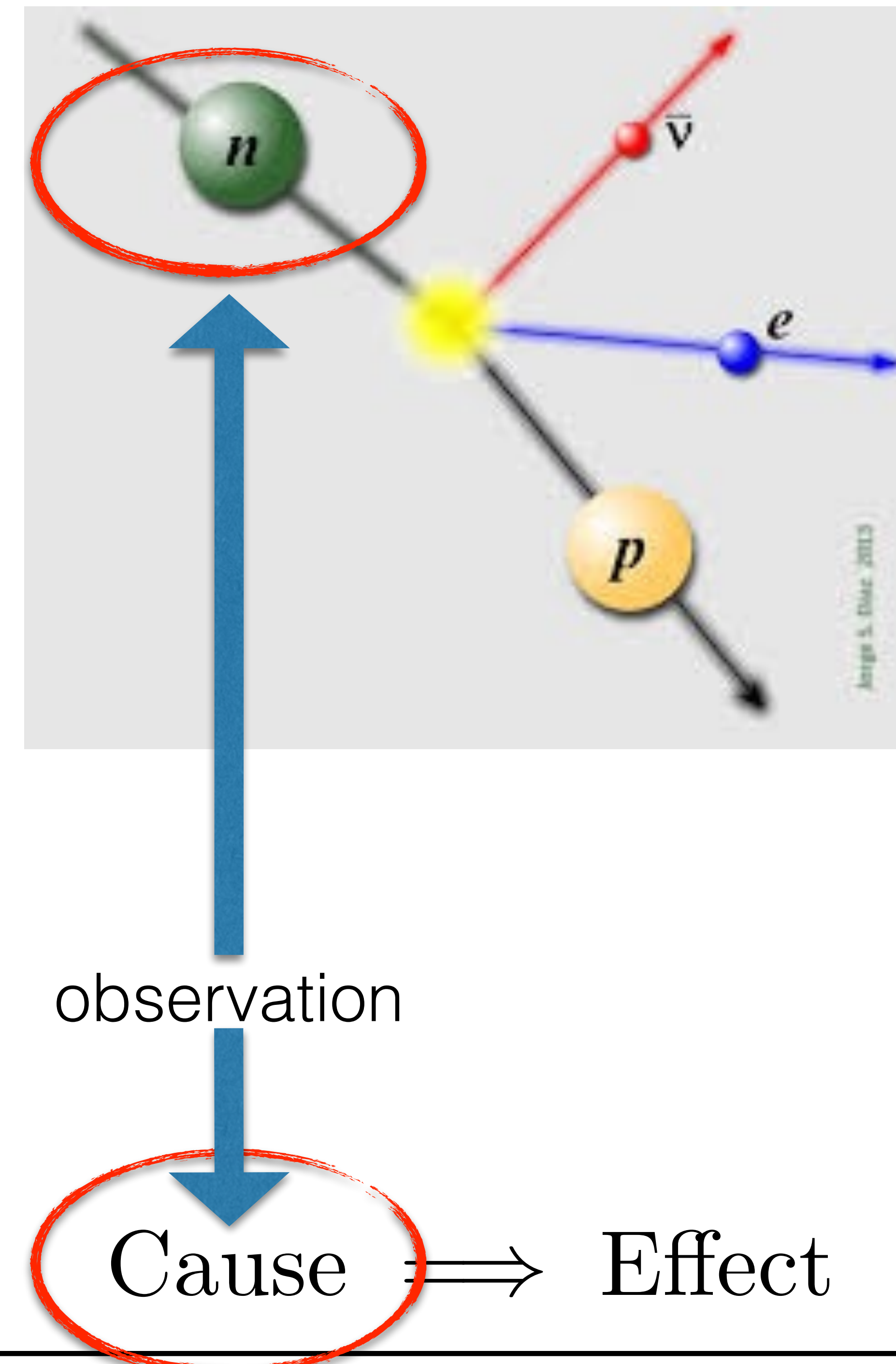
- Major premise (assumption):

$$n^0 \rightarrow p^+ + e^- + \bar{\nu}_e$$

- Minor premise (observation):
- particle x is a neutron $x \equiv n^0$
- Conclusion (inference):

$$x \rightarrow p^+ + e^- + \bar{\nu}_e$$

- Deductive inference is based on the certainty that a cause implies an effect and that one observes the cause



Inductive inference

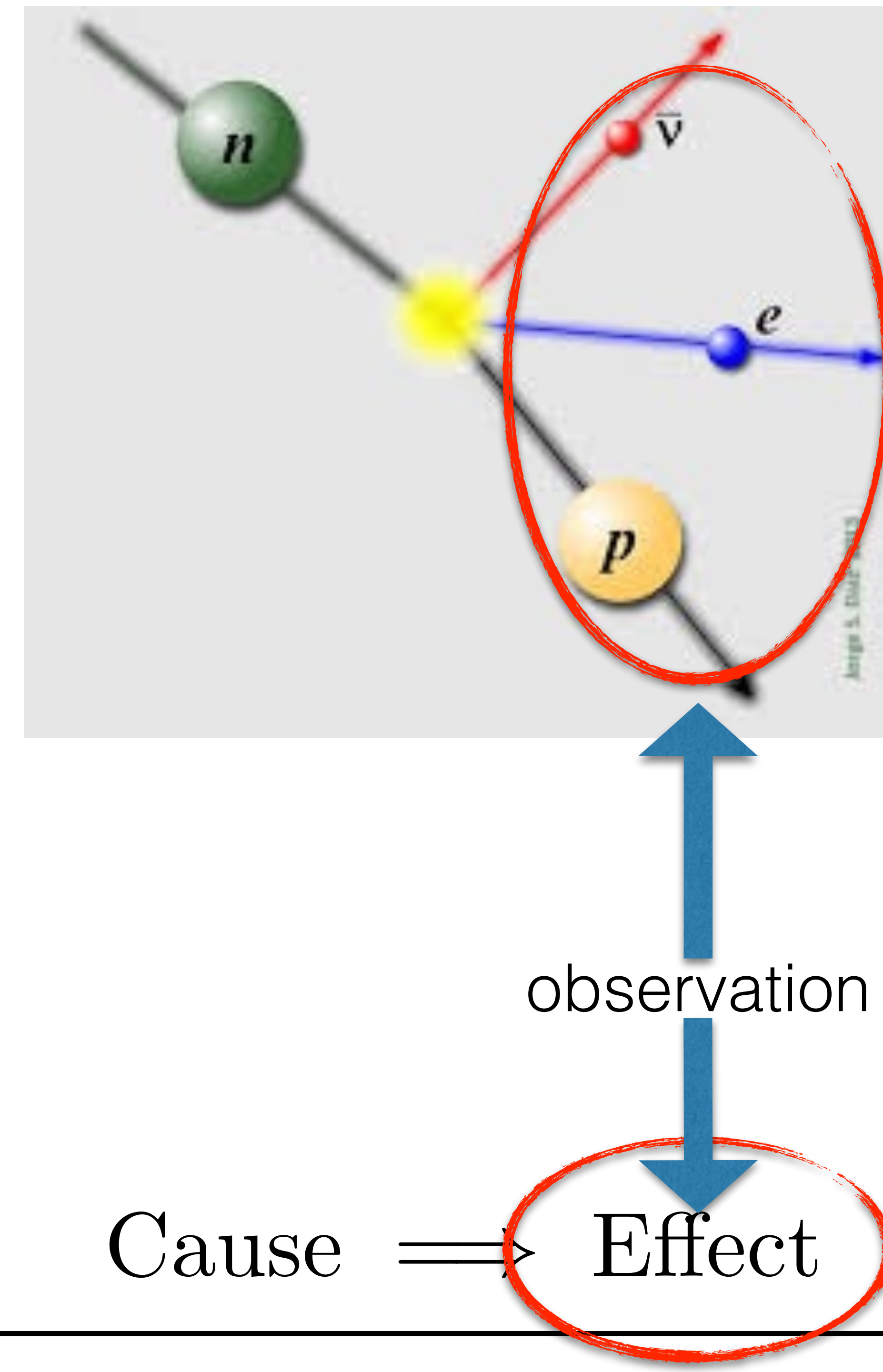
- Major premise (assumption):

$$n^0 \rightarrow p^+ + e^- + \bar{\nu}_e$$

- Minor premise (observation):

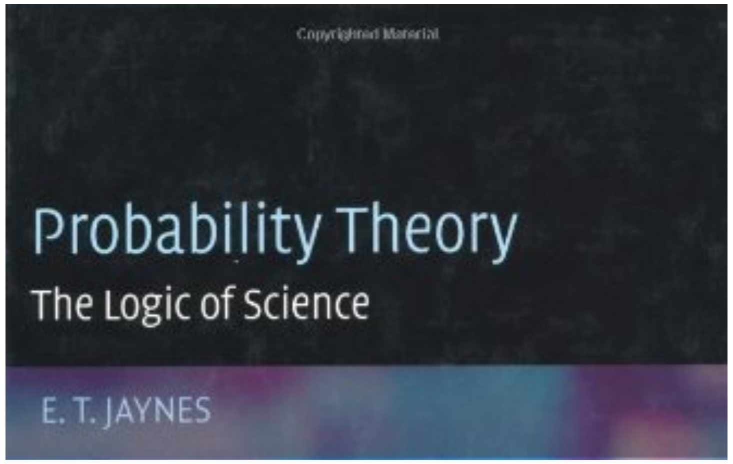
$$x \rightarrow p^+ + e^- + \bar{\nu}_e$$

- Conclusion (inference):
  - x is plausibly a neutron
- Inductive inference is based on observation of an effect and tries to reconstruct the cause



Bayesian theory: desiderata

- Bayesian theory is based on a set of “desiderata”:
    - Degrees of plausibility are represented by real numbers
    - Plausibility must be in qualitative agreement with rationality and common sense
    - Consistency:
        - if a conclusion can be reached in more than one way, every possible way must lead to the same result
        - the theory must account for all relevant information available
        - equivalent states of knowledge must lead to the same degree of plausibility assignment
- Probability is a map of degree of plausibility to [0,1]



Information based inference

- **Probabilities are always conditional**
  - Probability assignments depend on the "background information"

I: The die is unbiased

$$p(6|I) = \frac{1}{6}$$

I': The die is biased

$$p(6|I') > \frac{1}{6}$$

The quantitative rules

- Basic rules:
  - Product rule

$$p(AB|C) = p(A|BC)p(B|C)$$

  - Sum rule

$$p(A+B|C) = p(A|C) + p(B|C) - p(AB|C)$$
$$= p(A|C) + p(B|C) \iff p(AB|C) = 0$$

A and B are mutually exclusive (on the information C)

Bayes theorem

- From the product rule

$$p(AB|C) = p(A|BC)p(B|C)$$
$$= p(B|AC)p(A|C)$$

- Bayes theorem:

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)}$$

Bayes theorem

- Define the following statements:
  - A = H: hypothesis
  - B = D: observation
  - C = I: whatever relevant information

posterior, prior, likelihood, evidence

$$p(H|DI) = p(H|I)\frac{p(D|HI)}{p(D|I)}$$

$$p(D|I) = \sum_i p(H_i|I)p(D|H_iI)$$

marginalisation

Example

- To the (bat) board!

Model selection

- Consider 2 competing hypotheses (models)

$$H_1,\ H_2$$

- Given an observation D, which model is to be preferred?

$$p(H_1|DI) = p(H_1|I)\frac{p(D|H_1I)}{p(D|I)}$$

$$p(H_2|DI) = p(H_2|I)\frac{p(D|H_2I)}{p(D|I)}$$

Model selection

- The troublesome term

$$p(D|I) = \sum_i p(H_i|I)p(D|H_iI)$$

- Simplifies taking the ratio

$$O_{12} = \frac{p(H_1|I)p(D|H_1I)}{p(H_2|I)p(D|H_2I)} = \frac{p(H_1|I)}{p(H_2|I)} B_{12}$$

prior odds

Bayes' factor

UNIVERSITÀ DI PISA

Exhaustivity

- Note, the models might be exhaustive

$$\sum_j p(H_j|I) = 1$$

- But in general they are not

- Notable exception, "null test hypothesis"

$$p(H_1|I) + p(H_2|I) = 1 \implies H_1 = \overline{H}_2$$

Model selection

- If the model H_i depends on a set of parameters θ_i need to *marginalise* over them

$$p(D|H_i I) = \int_{\Theta} d\theta_i \, p(\theta_i|H_i I) p(D|\theta_i H_i I)$$

- Odds

$$O_{ij} = \frac{p(H_i|I)}{p(H_j|I)} \frac{\int_{\Theta_i} d\theta_i \, p(\theta_i|H_i I) p(D|\theta_i H_i I)}{\int_{\Theta_j} d\theta_j \, p(\theta_j|H_j I) p(D|\theta_j H_j I)}$$

Assigning probabilities

- **Indifference principle**: “if among the possible outcomes, there is no reason to prefer any of them over any other, then all outcomes should be equally probable.”

Assigning probabilities

- **Indifference principle**: "if among the possible outcomes, there is no reason to prefer any of them over any other, then all outcomes should be equally probable."

- **Invariance arguments**: $p(\sigma|I)d\sigma = p(\sigma'|I)d\sigma'$

$$\sigma' = k\sigma \iff p(\sigma|I) \propto \sigma^{-1}$$

Assigning probabilities

- **Indifference principle**: “if among the possible outcomes, there is no reason to prefer any of them over any other, then all outcomes should be equally probable.”

- **Invariance arguments**: $$p(\sigma|I)d\sigma = p(\sigma'|I)d\sigma'$$ $$\sigma' = k\sigma \iff p(\sigma|I) \propto \sigma^{-1}$$

- **Maximum entropy**: $$H(p) = -\sum_j p_j \log\left(\frac{p_j}{m_j}\right)$$

Indifference principle

- Take a 6-face die, if you have no reason to suspect the die is biased

$$\sum_{i=1}^{6} p_i = 1 \qquad p_i = p_j \ \forall i, j \qquad p_i = \frac{1}{6}$$

- In general, for N equally probable cases

$$p_i = \frac{1}{N}$$

Invariance

- I: we live in Friedmann-Robertson-Walker-LeMaitre Universe, how should I expect GW sources to be distributed (ignoring evolutionary effects)?

$$N = \int dV \, \frac{dN}{dV}$$

$$\frac{dN}{dV} = n_0$$

$$dV = D_L^2 \cos(\theta) dD_L d\theta d\phi; \quad z << 1$$

$$p(D_L, \theta, \phi | I) \propto D_L^2 \cos(\theta)$$

Uniform distribution

$$H(p) = -\sum_{j=1}^{N} p_j \log\left(\frac{p_j}{m_j}\right)$$

We want to maximise the entropy with some constraint

$$\sum_j p_j = 1$$

Only constraint

$$\delta\left[-\sum_{j=1}^{N} p_j \log\left(\frac{p_j}{m_j}\right) - \lambda\left(\sum_{j=1}^{N} p_j - 1\right)\right] = 0$$

Lagrange multiplier

$$\ldots$$

$$\lambda = -1$$

$$p_j = m_j$$

$$m_j = \frac{1}{N}$$

Gaussian distribution

- Similar calculation shows that if the constraint is

$$\sum_j p_j = 1$$

$$\sum_j (x_j - \mu)^2 p_j = \sigma^2$$

- The maximum entropy distribution is

$$p(x) = \frac{1}{\sqrt{2}\pi\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gaussian distribution

- Similar calculation shows that if the constraint is

$$\sum_j (x_j -$$

- The maxi

If the only known constraint is the variance, the Gaussian distribution is the least informative probability distribution.

$$p(x) = \frac{1}{\sqrt{2}\pi\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Assigning likelihoods

- Assume the hypothesis H depends on parameters θ
- The likelihood: $p(D|\theta H I)$
- Data $D = D_1,\ldots,D_n = d_1,\ldots,d_n$
- $H = X_1,\ldots,X_n$: datum di in $[x_i,x_i+dx_i]$
- $E = E_1,\ldots,E_n$: the error on d_i in $[e_i,e_i+de_i]$

Assigning likelihoods

- One can write $d_i = x_i + e_i$

$$p(X_i|\theta HI) \equiv f(x_i)$$

$$p(E_i|\theta HI) \equiv g(e_i)$$

- Consider joint distribution for D_i, E_i, X_i $\quad p(D_i E_i X_i|\theta HI)$

- Marginalise over E_i, X_i

$$p(D_i|\theta HI) = \int dE_i dX_i \, p(D_i E_i X_i|\theta HI)$$

$$= \int dE_i dX_i \, p(D_i|E_i X_i \theta HI) p(E_i|\theta HI) p(X_i|\theta HI)$$

Assigning likelihoods

$$d_i = x_i + e_i \implies p(D_i|E_i X_i \theta H I) = \delta(d_i - x_i - e_i)$$

$$p(D_i|\theta H I) = \int dx_i f(x_i) \int de_i g(e_i) \delta(d_i - x_i - e_i)$$

$$= \int dx_i f(x_i) g(d_i - x_i)$$

- Deterministic models $f(x_i) = \delta(x_i - m(x_i;\theta))$

$$p(D_i|\theta H I) = g(d_i - m(x_i;\theta))$$

- If the errors are independent

$$p(D|\theta H I) = \prod_{i=1}^{N} g(d_i - m(x_i;\theta))$$

Assigning likelihoods

$$d_i = x_i + e_i \implies p(D_i|E_i X_i \theta H I) = \delta(d_i - x_i - e_i)$$

$$p(D_i|\theta H I) = \int dx_i f(x_i) \int de_i g(e_i) \delta(d_i - x_i - e_i)$$

$$= \int dx_i f(x_i) g(d_i - x_i)$$

- Deterministic models $f(x_i) = \delta(x_i - m(x_i;\theta))$

$$p(D_i|\theta H I) = g(d_i - m(x_i;\theta))$$

- If the errors are independent

$$p(D|\theta H I) = \prod_{i=1}^{N} g(d_i - m(x_i;\theta))$$

The noise probability distribution defines the likelihood function

Part 2

Stochastic processes

Stochastic processes

- Stochastic process is a family of random variables

$$\{X(t)\ \ t \in T\}$$

- We will deal with discrete time processes where the set of time samples is potentially infinite, but countable

$$\{x(t_1), \ldots, x(t_n)\} \equiv \{x_1, \ldots, x_n\}$$

- The state space (the set of possible realisations of the process) is infinite

Process properties

- Process is characterised by its statistics
  - Mean: $< n(t) >= \mu(t)$
  - Autocovariance: $< n(t)n(t') >= C(t,t')$
- If the statistics do not depend on time we have a stationary process
  - e.g. white noise Gaussian process

Wide-sense stationarity

- For a wide sense stationary process

$$< n(t) > = \mu(t) = \mu(t+\tau)$$
$$< n(t)n(t') > = C(t-t',0)$$

- Constant mean
- Covariance function depends only on the time difference

Detector output process

- GW detector output is modelled as a wide-sense stationary stochastic process
  - Given a set of (regularly spaced) times $\{t_1, \ldots, t_n\}$
  - The detector registers $\{d_1, \ldots, d_n\}$
- In the absence of a GW signal

$$d_i = n_i \ \forall i$$

- When (if) a signal is present

$$d_i = n_i + h_i(\theta)$$

$$h_i(\theta) = h(t_i, \theta)$$

see also Finn, arXiv:9209010

Noise distribution

- Assume that the only non-null moment of the probability distribution for the process realisation is the second (the autocovariance)

- Using the maximum entropy principle

$$p(\vec{n}|I) \propto e^{-\frac{1}{2}\vec{n}^t C^{-1}\vec{n}}, \quad \vec{n} = n(t_1), \ldots, n(t_K)$$

$$< \vec{n} >= 0$$

The noise model

- The matrix C is the covariance matrix

$$C = \begin{pmatrix} C(t_1,t_1) & C(t_1,t_2) & \cdots & C(t_1,t_K) \\ C(t_2,t_1) & C(t_2,t_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ C(t_K,t_1) & \cdots & \cdots & C(t_K,t_K) \end{pmatrix}$$

$$C(t_i,t_j) = < n(t_i) n(t_j) >$$

The noise model

- In general, the auto-covariance function depends on 2 time values (non-stationary process)

- However, we assumed that $n(t)$ is a wide-sense stationary process

$$C(t_i, t_j) \equiv C(t_i - t_j, 0) \equiv C(\tau) \equiv < n(t)n(t+\tau) >$$

- Assume now that we are sampling the process $n(t)$ at regular intervals

$$t_j = t_0 + j\Delta t$$

The noise model

- The covariance matrix becomes a Toeplitz matrix

$$\begin{pmatrix} C(0) & C(\Delta t) & \cdots & C(K\Delta t) \\ C(-\Delta t) & C(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ C(-K\Delta t) & \cdots & \cdots & C(0) \end{pmatrix}$$

- Toeplitz matrices are asymptotically equivalent to circulant matrices
    - Diagonalised by same base in the limit $K \to \infty$

Circulant matrices

- Circulant matrices have the general form

$$
\begin{pmatrix}
a_0 & a_{n-1} & \cdots & a_1 \\
a_1 & a_0 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
a_{n-1} & \cdots & \cdots & a_0
\end{pmatrix}
$$

- All circulant matrices have the same eigenvectors and eigenvalues

Eigenvalues & eigenvectors

- They are:

$$v_j = \frac{1}{\sqrt{n}}(1, \omega_j, \omega_j^2, \ldots, \omega_j^{n-1})^t$$

$$\omega_j = e^{-\frac{2ij\pi}{n}}$$

$$\lambda_j = \sum_{k=0}^{n-1} a_k e^{-2ij\pi\frac{k}{n}}$$

- Which are nothing more than the Discrete Fourier Transform series

Fourier Matrix

- The DFT matrix

$$F = \begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{K-1} \\ \vdots & \omega^2 & \omega^4 & \cdots & \omega^{2(K-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{K-1} & \omega^{2(K-1)} & \cdots & \omega^{(K-1)(K-1)} \end{pmatrix}$$

$$F \cdot F^{-1} = F^{-1} \cdot F = I \qquad F^{-1} = F^{\dagger}$$

$$\det(F) = 1$$

- diagonalises any circulant matrix and thus any wide-sense stationary process covariance matrix

Diagonalised circulant matrix

- It thus follows that any circulant matrix is diagonalised as

$$F^{-1}AF = \mathrm{Diag}(\sum_{j=0}^{K-1} a_j e^{-2\pi i n \frac{j}{K}})$$

- It follows that the same is true for a wide-sense stationary covariance matrix

- The elements of a diagonalised covariance matrix are the DFT of the autocorrelation function

The noise distribution

- We have then:

$$\vec{n}^t C^{-1} \vec{n} = \vec{n}^t (F^\dagger F) C^{-1} (F^\dagger F) \vec{n} =$$

$$= (F\vec{n})^\dagger (F C^{-1} F^\dagger)(F\vec{n}) =$$

$$= \tilde{\vec{n}}^t S^{-1} \tilde{\vec{n}}$$

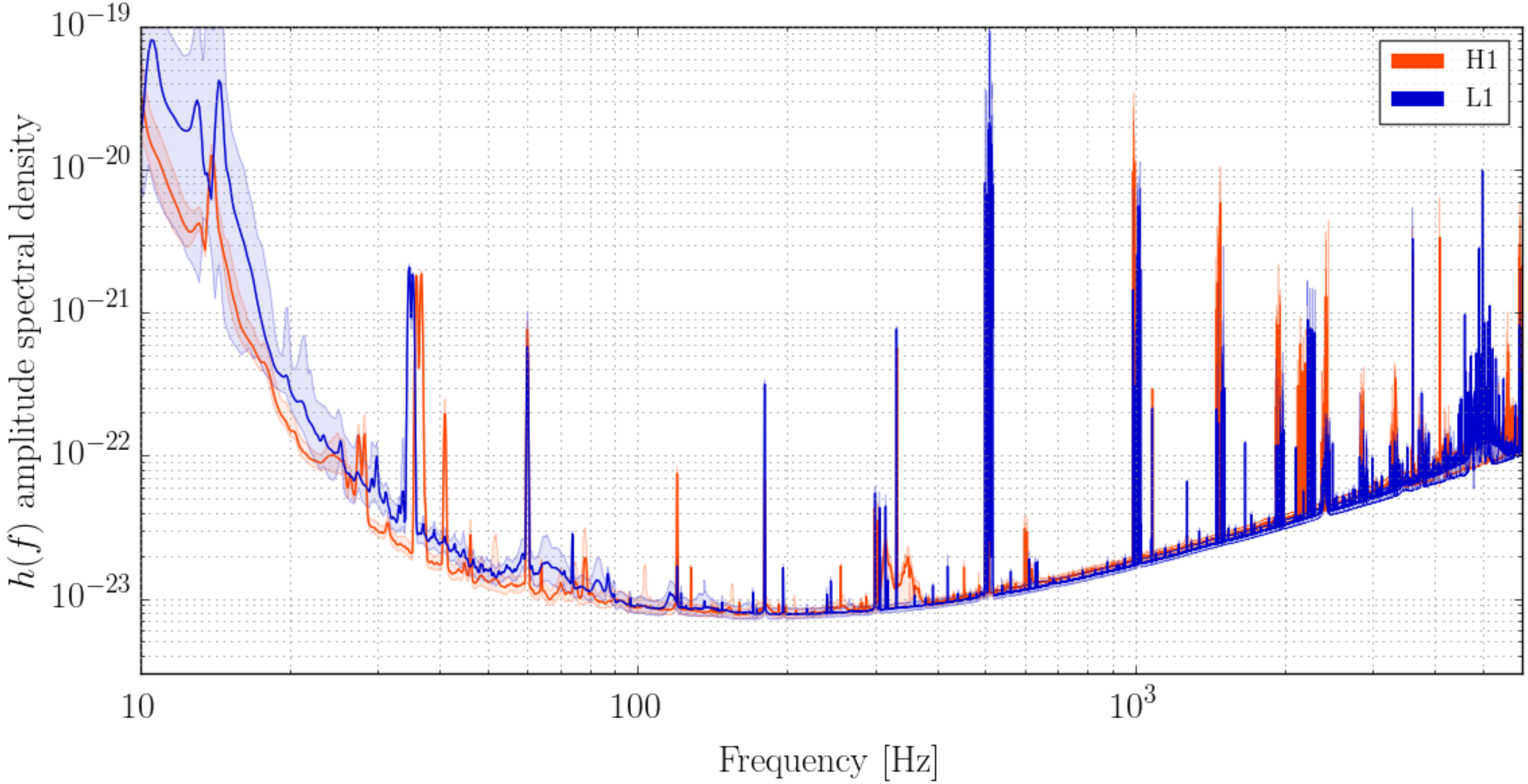
- And $S_n = \sum_{j=0}^{K-1} C_{nj} e^{-2\pi i n \frac{j}{K}}$

- Parseval's Theorem

$$< n(t)n(t+\tau) > = C(\tau)$$

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

Power spectral density



Matched filtering

- Detector output $d = h + n$
- Filter $A(t) = \int d\tau F(t+\tau)a(\tau) = \int df \tilde{F}^*(f)\tilde{a}(f)$
- Filtered output

$$\int df \tilde{F}^*(f)\tilde{d}(f) = \int df \tilde{F}^*(f)\tilde{h}(f) + \int df \tilde{F}^*(f)\tilde{n}(f)$$

$$D = H + N$$

- signal-to-noise ratio

$$SNR \equiv \rho = \frac{H^2}{< N^2 >}$$

Matched filtering

$$< N^2 > = \int df |\tilde{F}(f)|^2 S_n(f)$$

$$|H|^2 = \left| \int df \tilde{F}^*(f) \tilde{h}(f) \right|^2$$

- Multiply and divide numerator by $\sqrt{S_n(f)}$ and use the Schwarz inequality

- Optimal filter

$$\tilde{F}^*(f) = C \frac{\tilde{h}(f)}{S_n(f)}$$

Signal-to-noise ratio

- Use optimal filter to define a scalar product

$$(a|b) = 4\mathrm{Re}\int df\,\frac{\tilde{a}^*\tilde{b} + \tilde{a}\tilde{b}^*}{S_n(f)}$$

- matched filter SNR $\rho^2 = (d|h)$

- optimal SNR $\rho^2_{opt} = (h|h)$

- Frequency domain Likelihood

$$p(D|\theta HI) = e^{-\frac{(d-h|d-h)}{2}}$$

Multiple detectors

- Easily generalised to several detectors

$$p(D_1, \ldots, D_k|\theta HI) = \prod_j p(D_j|\theta HI)$$

- Network SNRs

$$\rho^2 = \sum_j (d_j|h_j)$$

$$\rho^2_{opt} = \sum_j (h_j|h_j)$$

Match

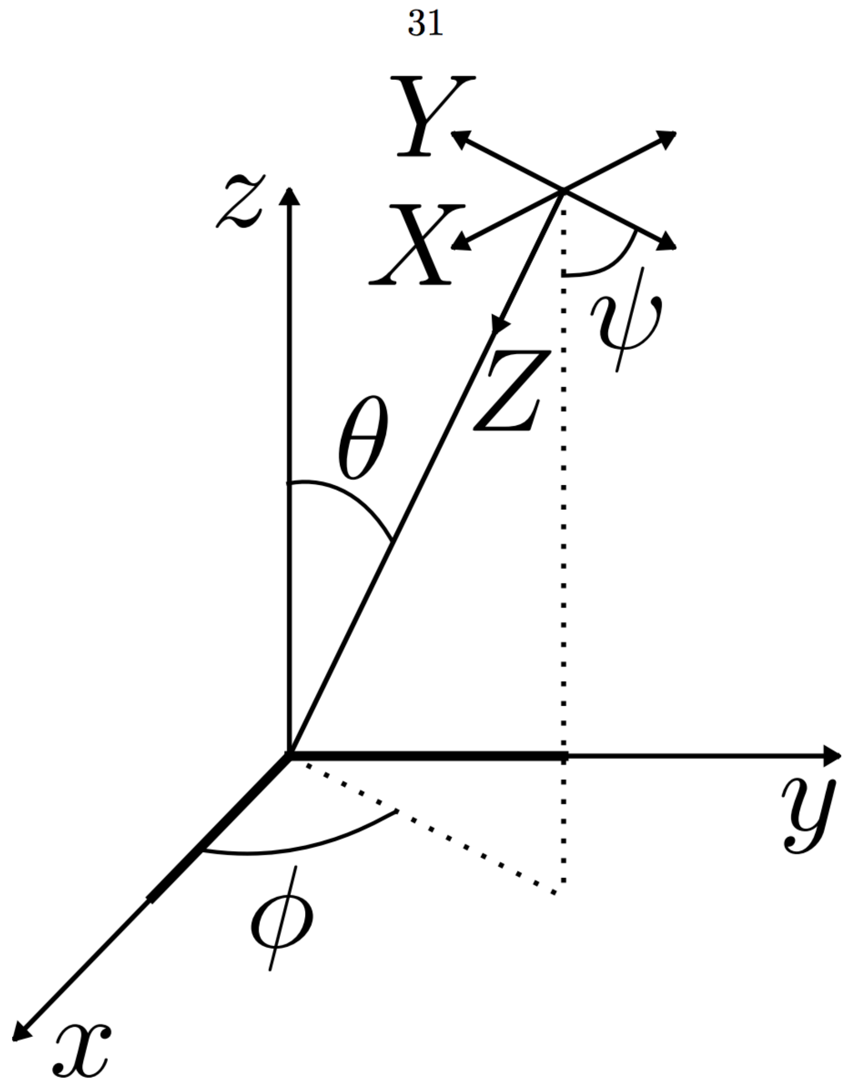
- Distance in the waveform manifold

$$M = \frac{(h_1(\theta)|h_2(\theta))}{\sqrt{(h_1(\theta)|h_1(\theta))(h_2(\theta)|h_2(\theta))}}$$

- Maximise over time and phase: faithfulness
- Maximise over θ : fitting factor

$$\rho = M\rho_t \implies f_{lost} = 1 - M^3$$

Signal projection



$$F_+ = \frac{1}{2}\left(1+\cos^2\theta\right)\cos 2\phi\cos 2\psi + \cos\theta\sin 2\phi\sin 2\psi$$

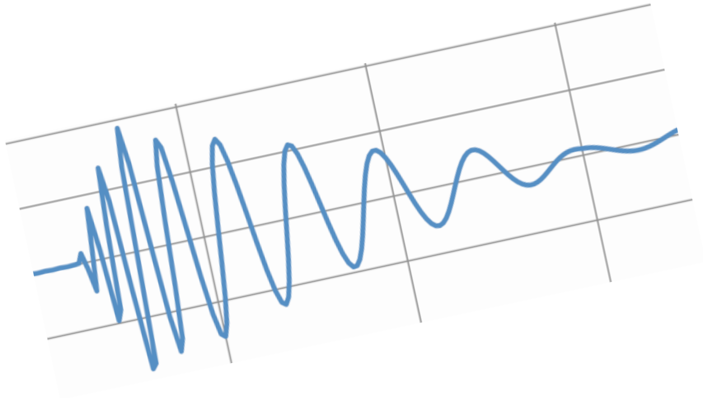
$$F_\times = \frac{1}{2}\left(1+\cos^2\theta\right)\cos 2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi$$

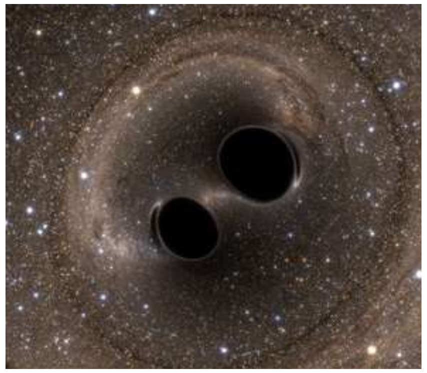
Schutz:arXiv:1102.5421

Signal projection

- Each detector sees a different signal

$$h = e^{2\pi i \vec{r}\cdot\vec{n} f}\left[F_+ h_+(\theta, f) + F_\times h_\times(\theta, f)\right]$$

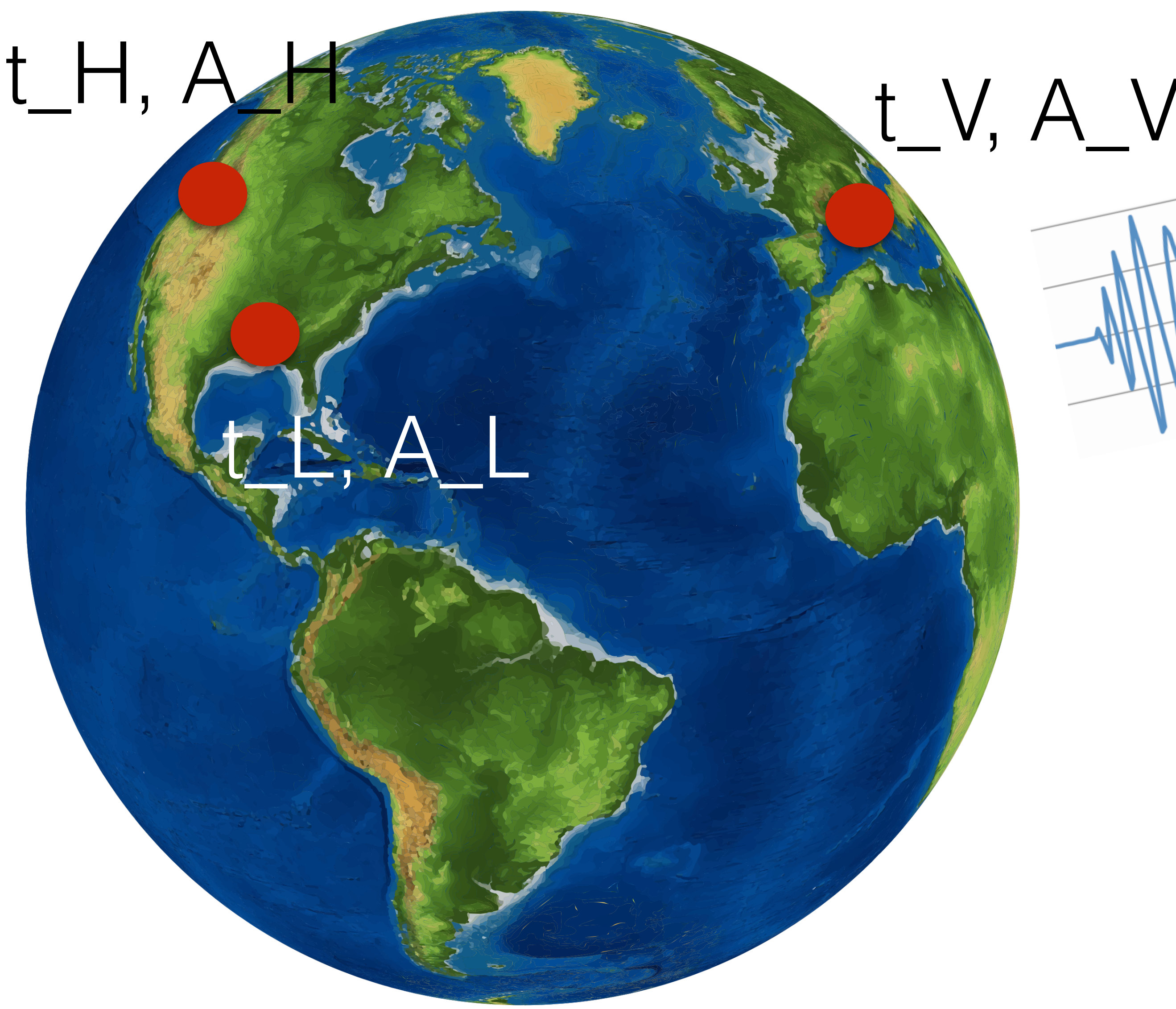


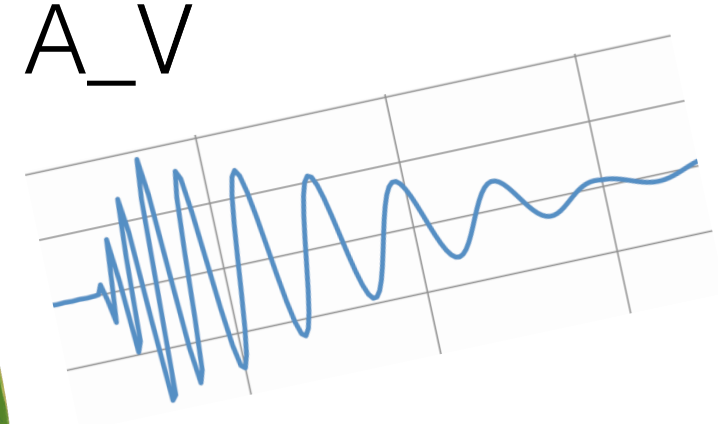


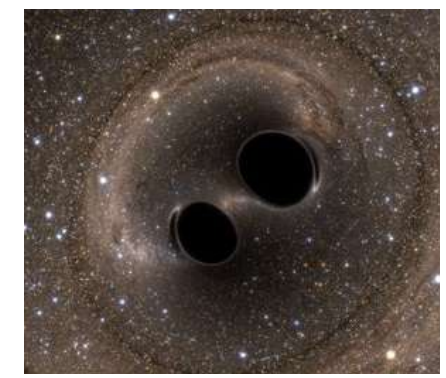
Signal projection

- Each detector sees a different signal

$$h = e^{2\pi i \vec{r}\cdot\vec{n} f}\left[F_+ h_+(\theta, f) + F_\times h_\times(\theta, f)\right]$$



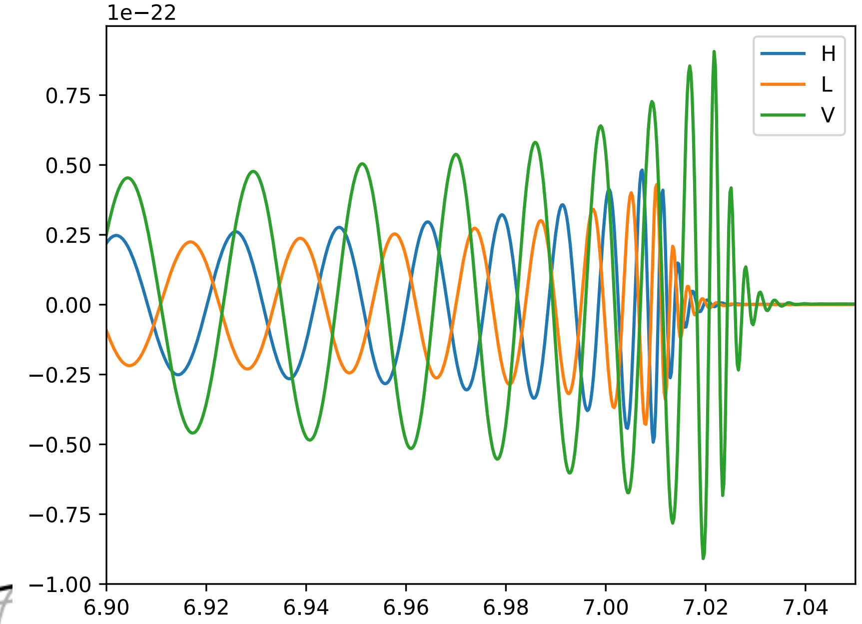


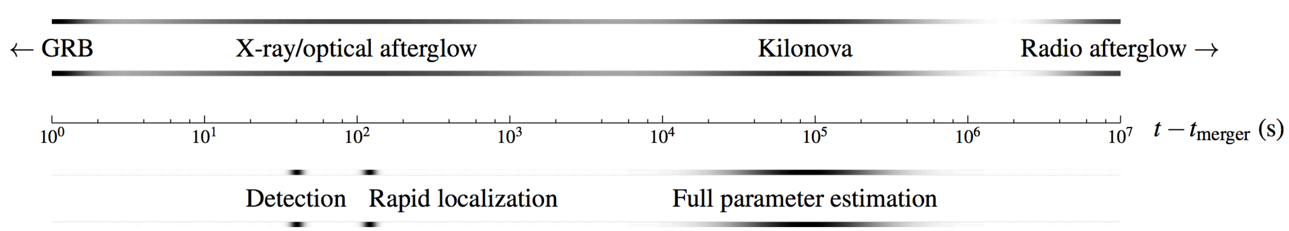


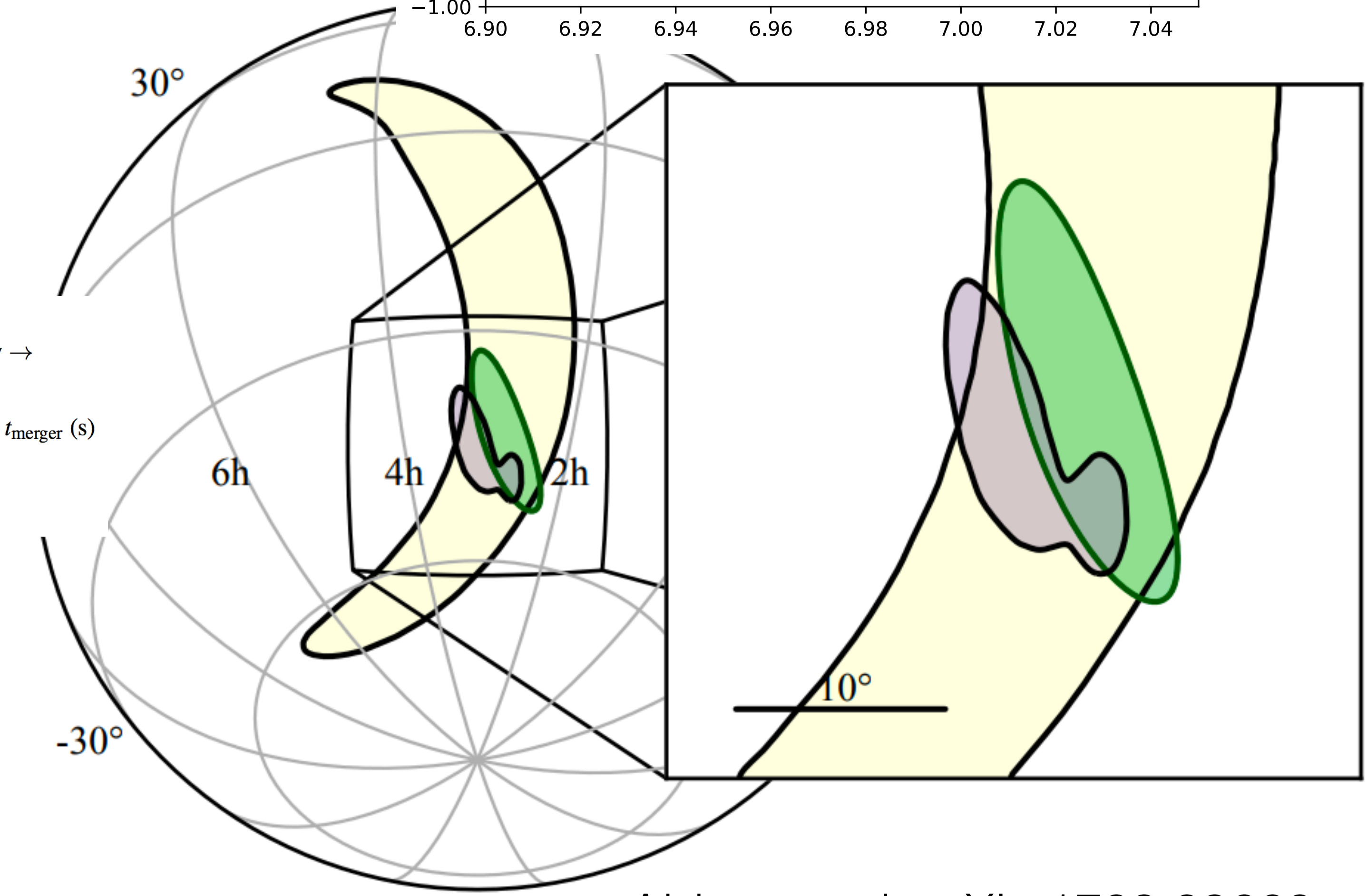
UNIVERSITÀ DI PISA

Quick localisation

- Relative amplitudes and time differences allow for rapid localisation O(100) s







Singer at al, arXiv:1404.5623

Abbott at al, arXiv:1709.09660

Part 3

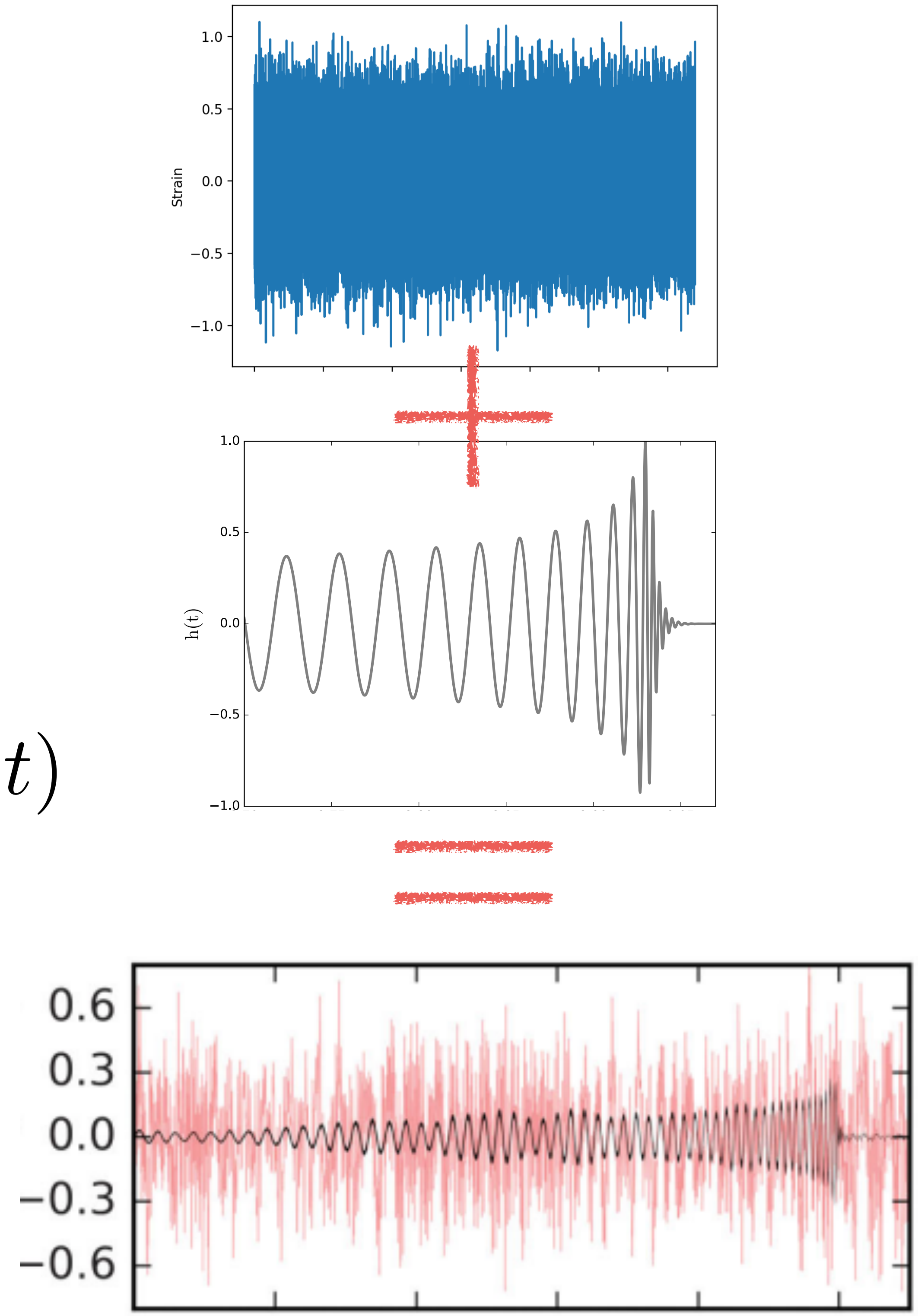
Detection of GW

Detector model

- The detector output is linear

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

- where $h(t;\theta)$ is the gravitational wave strain and $n(t)$ is the noise time series



UNIVERSITÀ DI PISA

Detection of GW

- Compute odds ratio for the hypotheses
  - S: (Unknown) signal plus noise
  - N: pure noise

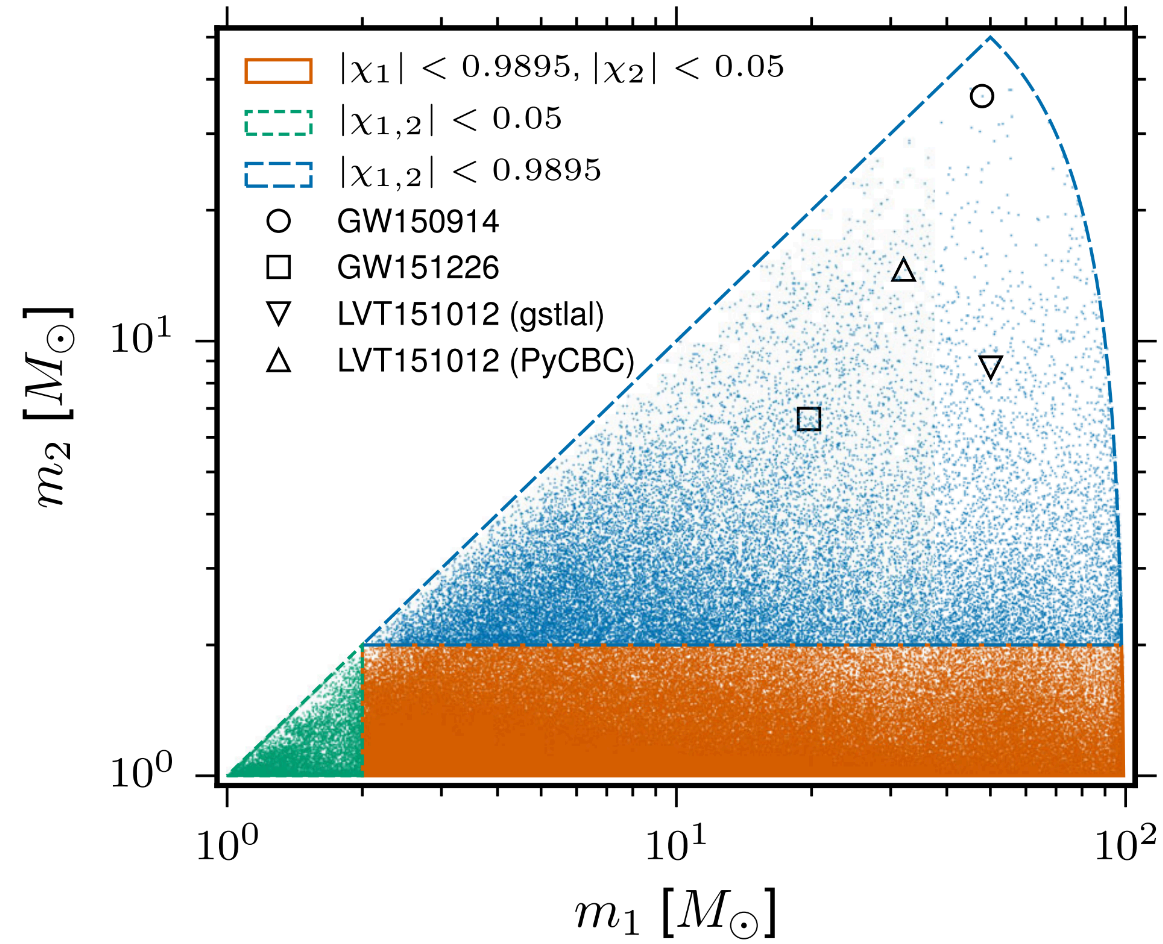
$$O_{SN} = \frac{p(S|I)}{p(N|I)} \frac{\int_\Theta d\theta \, p(\theta|SI)p(D|\theta SI)}{p(D|NI)}$$

- Computationally unfeasible, yet
- Instead of marginalising, maximise the likelihood

$$p(D|\theta SI) = e^{-\frac{(d-h|d-h)}{2}}$$

Template banks

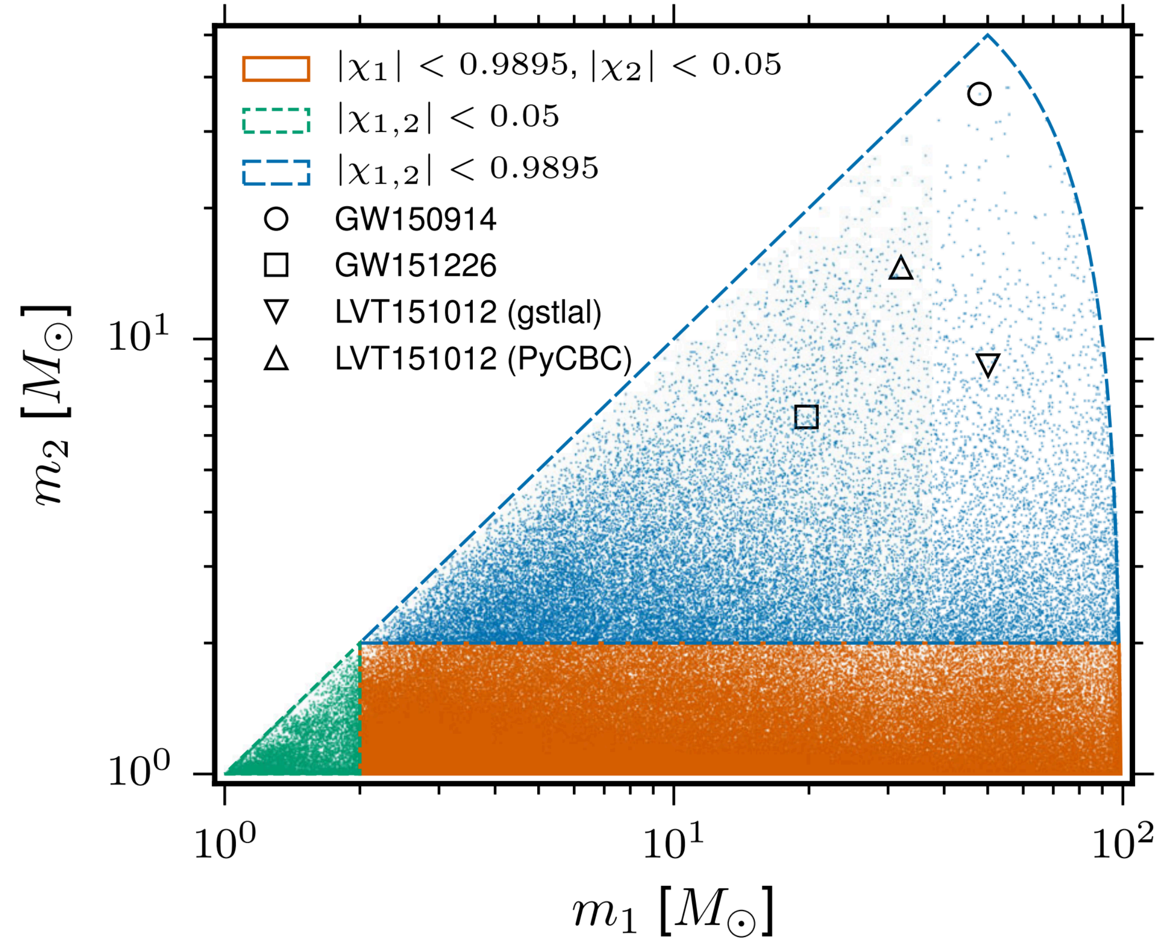
- Construct template banks O(200000) templates
- Template distribution
  - Following some intrinsic metric
  - Stochastically
- Spacing between templates such that



$$\max(1-M) \leq 0.03 \implies \text{less than 10\% loss in event rate}$$

Filtering

- Data are filtered against banks of O(200000) templates
- Compute SNR for each template
- Compute sanity check statistics (energy distribution per frequency octave)
- Rank the templates according to the detection statistics



$$\hat{\rho} = \begin{cases} \rho \Big/ [(1 + (\chi_r^2)^3)/2]^{\frac{1}{6}}, & \text{if } \chi_r^2 > 1, \\ \rho, & \text{if } \chi_r^2 \leq 1. \end{cases}$$

$$\chi^2 = p \sum_{i=1}^{p} \left[\left(\frac{\rho_{\cos}^2}{p} - \rho_{\cos,i}^2 \right)^2 + \left(\frac{\rho_{\sin}^2}{p} - \rho_{\sin,i}^2 \right)^2 \right]$$

$$\chi_r^2 = \frac{\chi^2}{(2p-2)}$$

Significance estimation

- Significance of detection is evaluated against the hypothesis that noise could give raise to random trigger

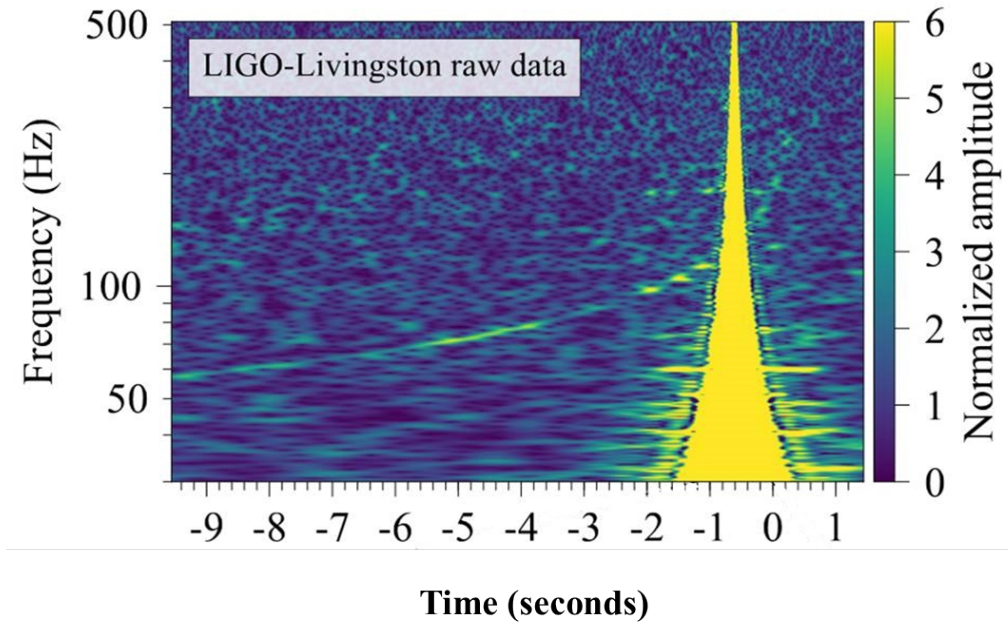
- GW150914: 5.3 sigma, FAR < 6×10^{-7} yr^{-1}

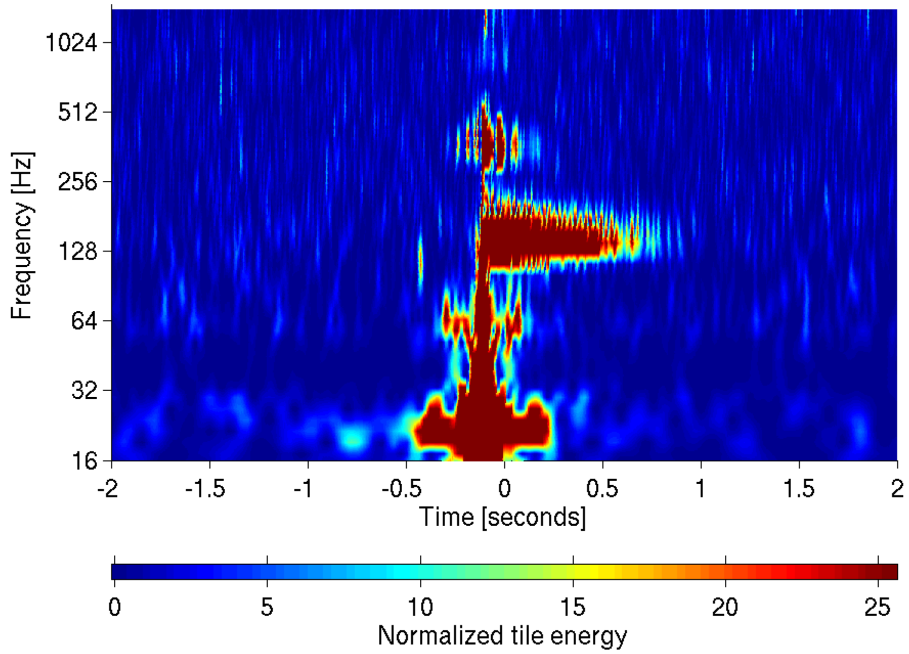
- O1 had 51.5 days of data. How is it possible?

- Timeslides

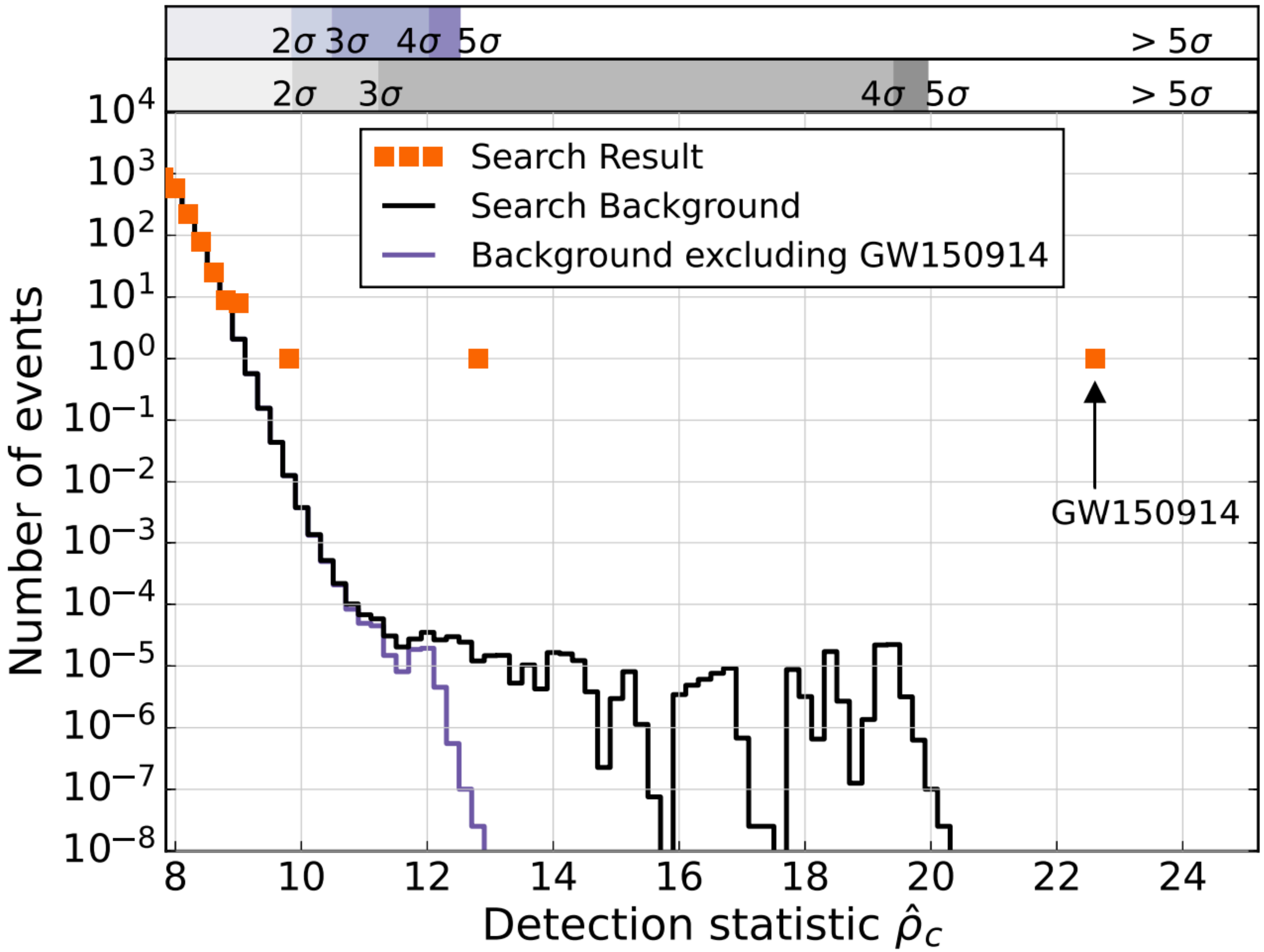
Timeslides

- Pure Gaussian noise $p(\rho|I) \propto e^{-\frac{1}{2}\rho^2}$
- The noise is not exactly Gaussian (glitches)
- Construct the empirical distribution of your detection statistics
- Chop the data in chunks and permute them in time
    - Synthetic, incoherent noise realisations
    - Background distribution





Significance

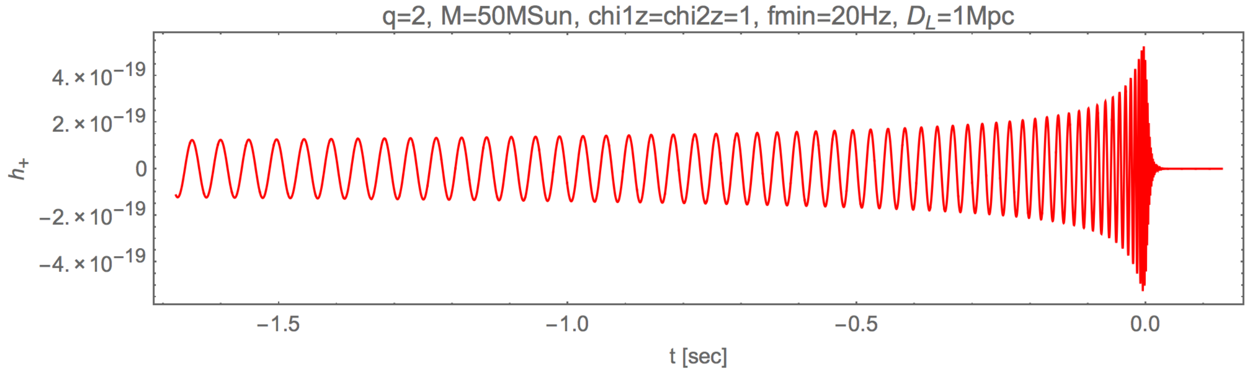


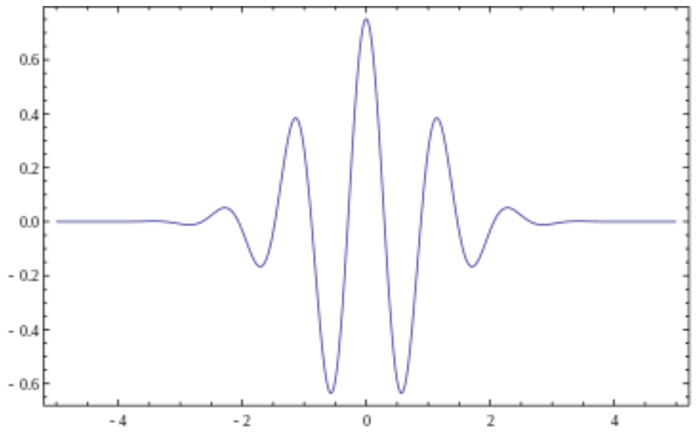
Part 4

Inference of parameters

Signal model

- Signal model is any deterministic signal which is not noise
    - Accurate prediction (e.g. EOB)
    - Superposition of some basis function (wavelets)
- Signal depends on a set of parameters
    - physical for GR models
    - coefficients of projection on basis for unmodelled





Parameter estimation

- $h(t;\theta)$ depends on a set of parameters θ
  - D=9 for non-spinning binaries: masses, orientation, sky location, reference time and phase, luminosity distance
  - D=15 in general: spin vectors
  - More parameters for extra physics (e.g. BH charges, tests of GR, tidal effects, etc…)

Parameter estimation

- Parameters are estimated computing the posterior distribution for all of them

  - joint distribution

$$p(\theta|DSI) = \frac{p(\theta|SI)p(D|\theta SI)}{\int_{\Theta} p(\theta|SI)p(D|\theta SI)}$$

Large dimensional integral
numerical methods

Assigning confidence

- In addition to posteriors, one can compute the odds between the noise and signal + noise hypotheses

$$O_{SN} = \frac{p(S|I)}{p(N|I)} \frac{\int_{\Theta} d\theta \, p(\theta|SI) p(D|\theta SI)}{p(D|NI)}$$

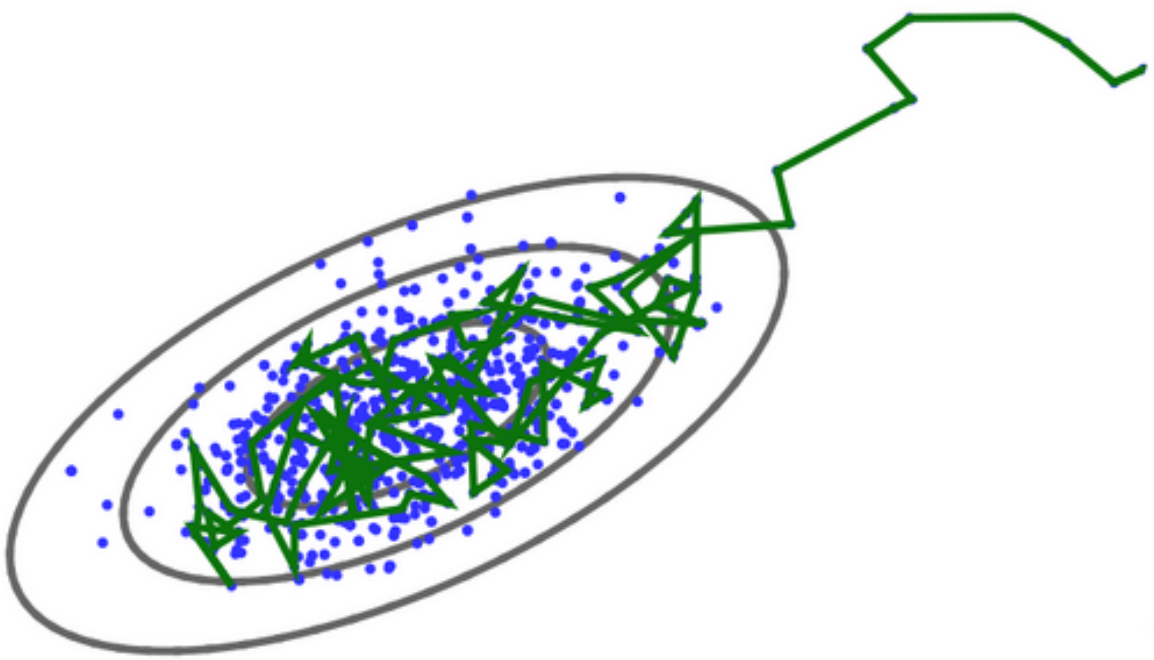
Bayes' factor

Numerical implementations

- The LVKC parameter estimation algorithms and necessary infrastructure are implemented in the LALInference package as part of the LIGO Algorithm Library (LAL)
  - Parallel tempering Markov Chain Monte Carlo
  - Nested Sampling

MCMC in a nutshell

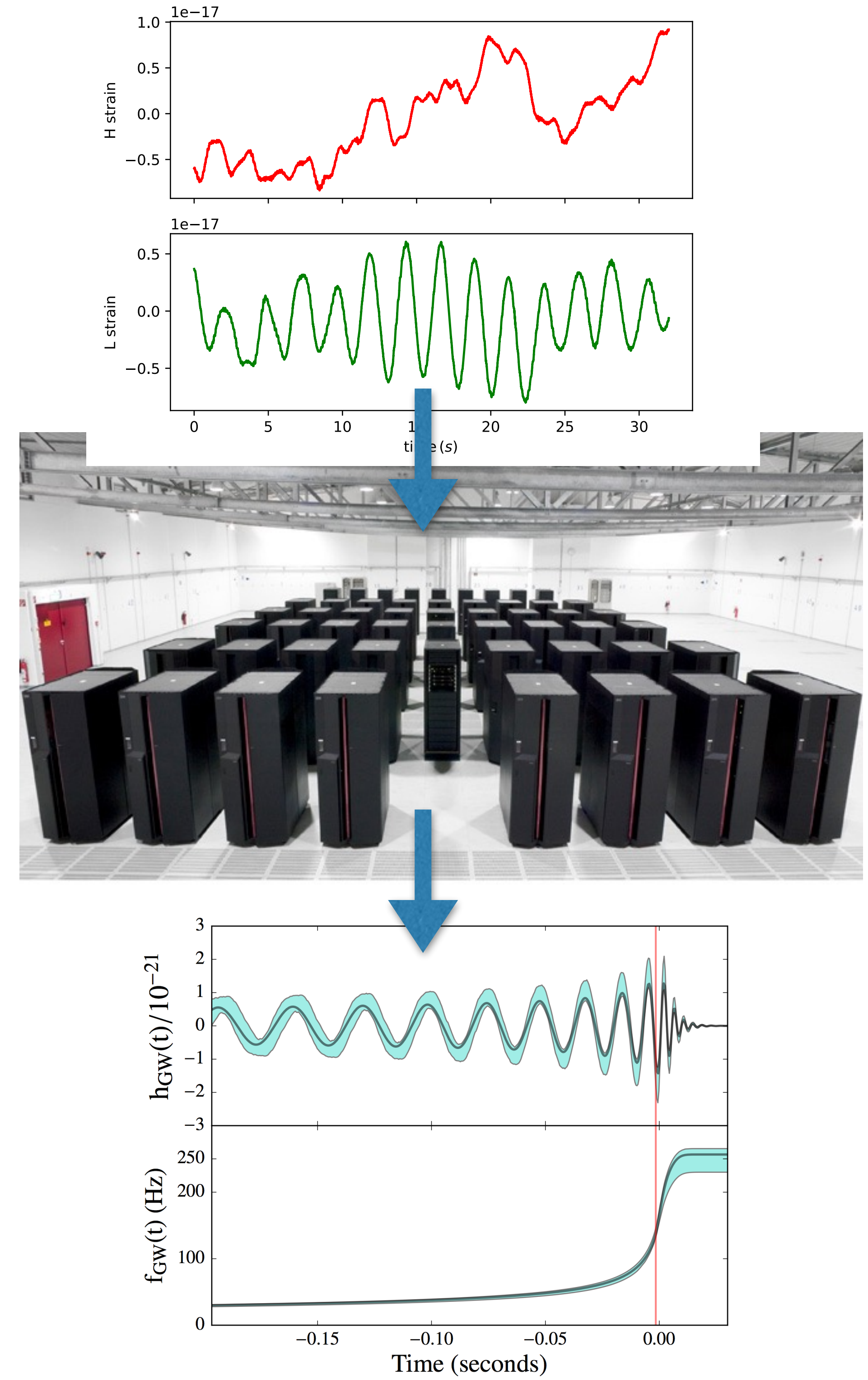
- while you are not done:
  - pick a value of the parameters (sample) as prescribed
  - predict the waveform
  - subtract it from the data
  - if it looks like noise:
    - keep it and pick the next one "around" it
  - else:
    - pick the next one somewhere else



UNIVERSITÀ DI PISA

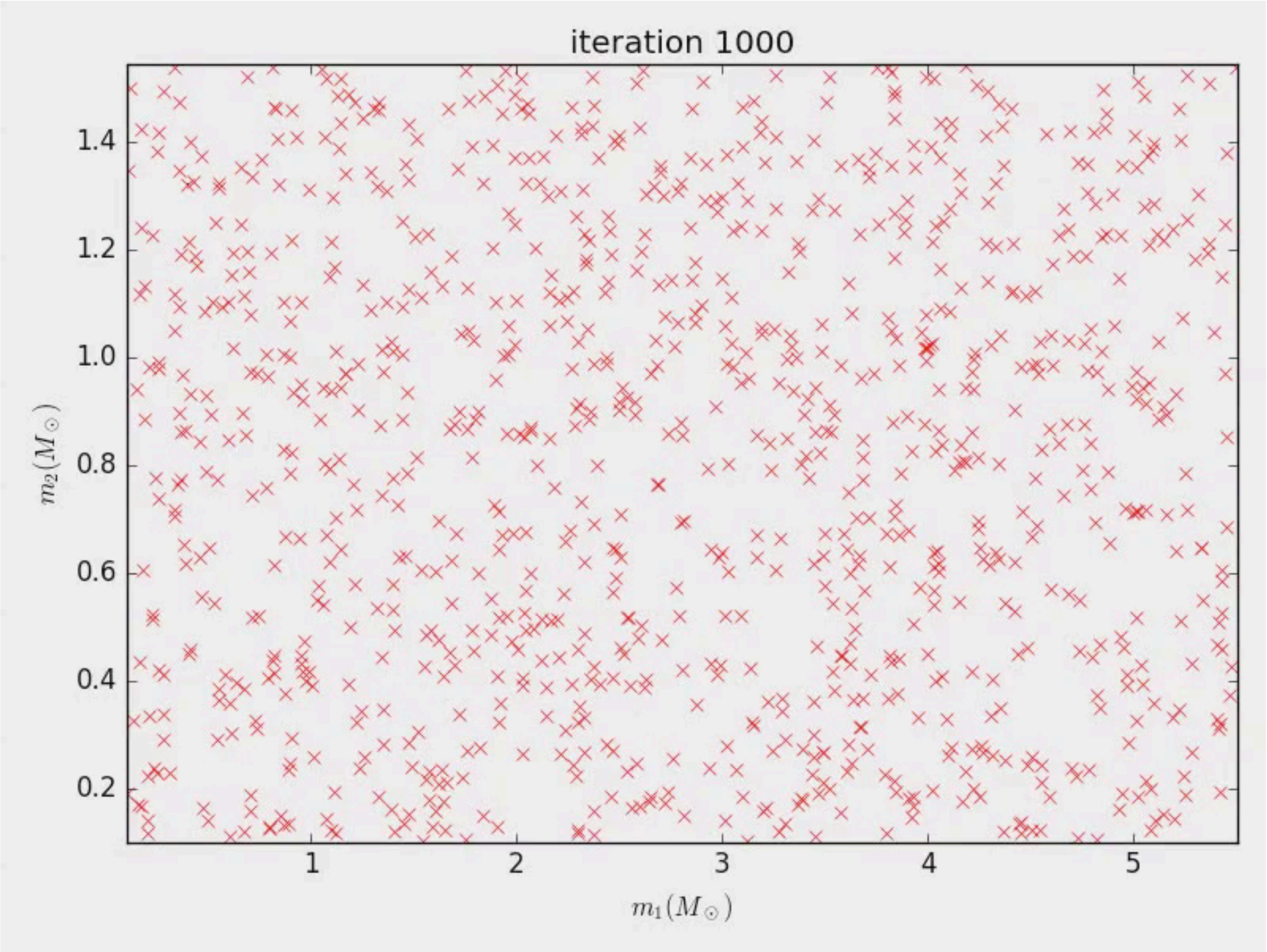
Some numbers

- Parameter estimation is an expensive business
  - Generation of O(10^7) templates
  - If a template takes 1s to generate $\Longrightarrow \quad \tau \sim \mathrm{months}$
  - Need for fast AND accurate waveform models
- EOBNR(ROM) & Phenom



Exploring the parameter space

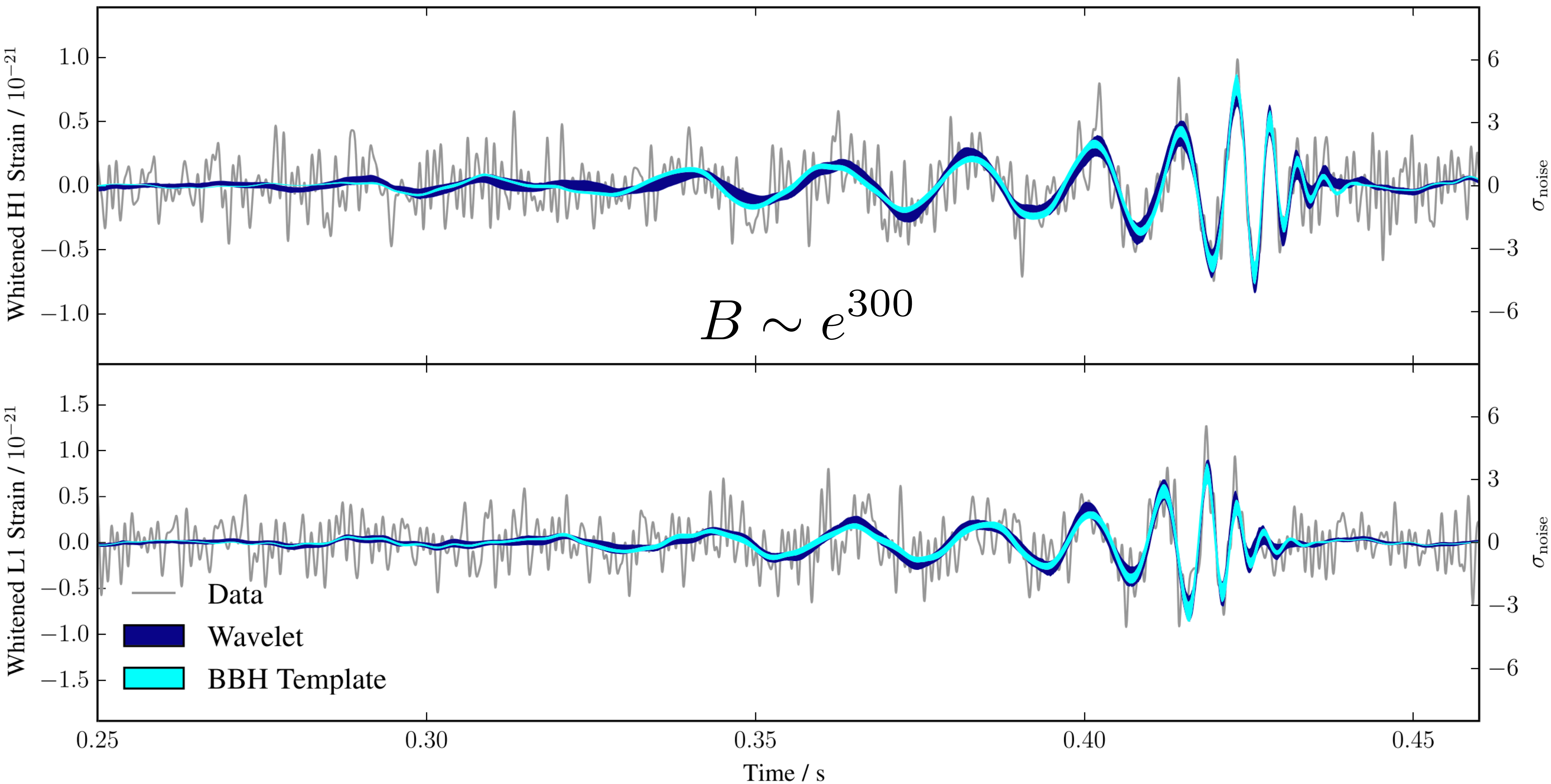
- nested sampling algorithm



Part 5

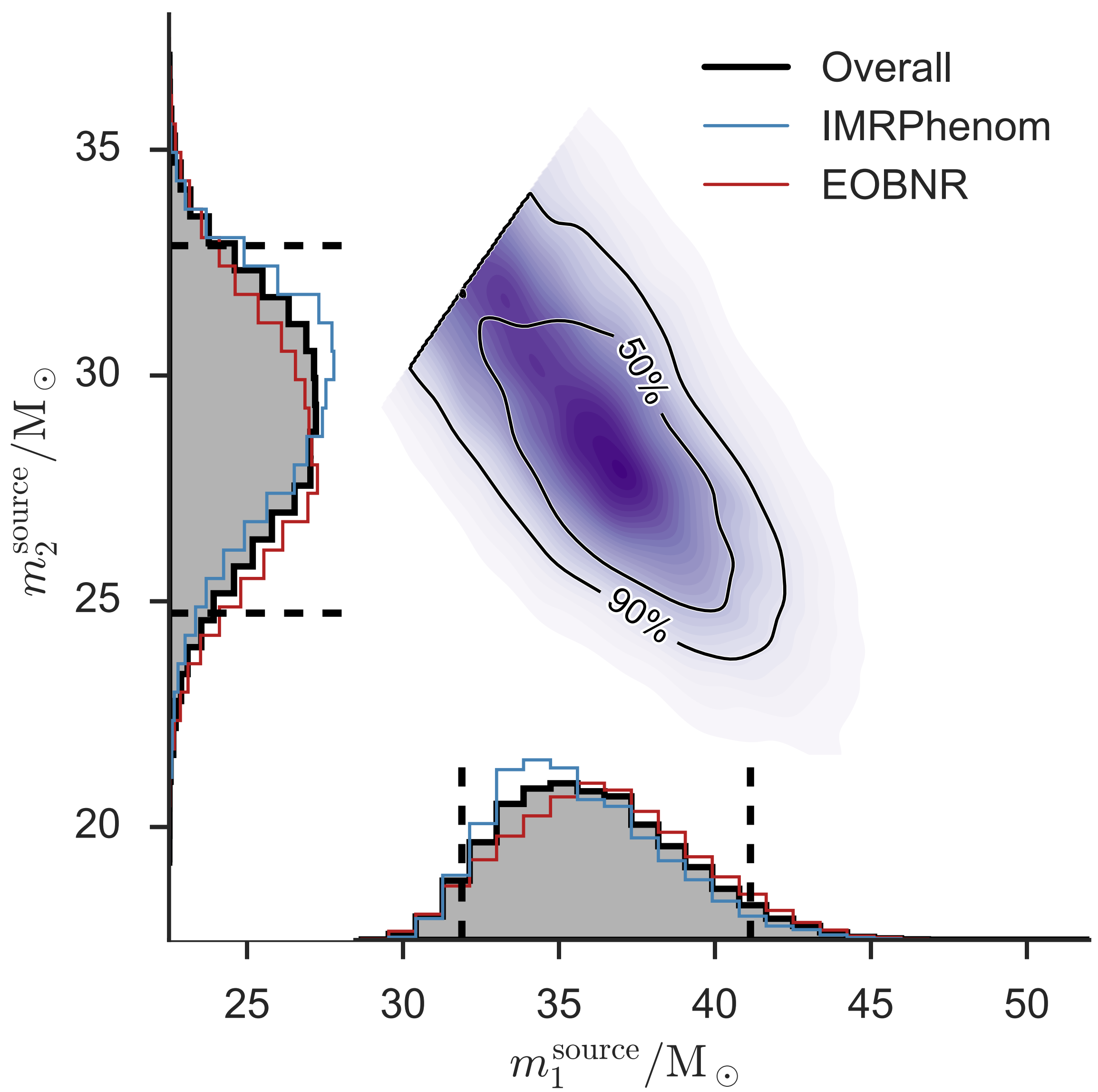
Selected posterior distributions

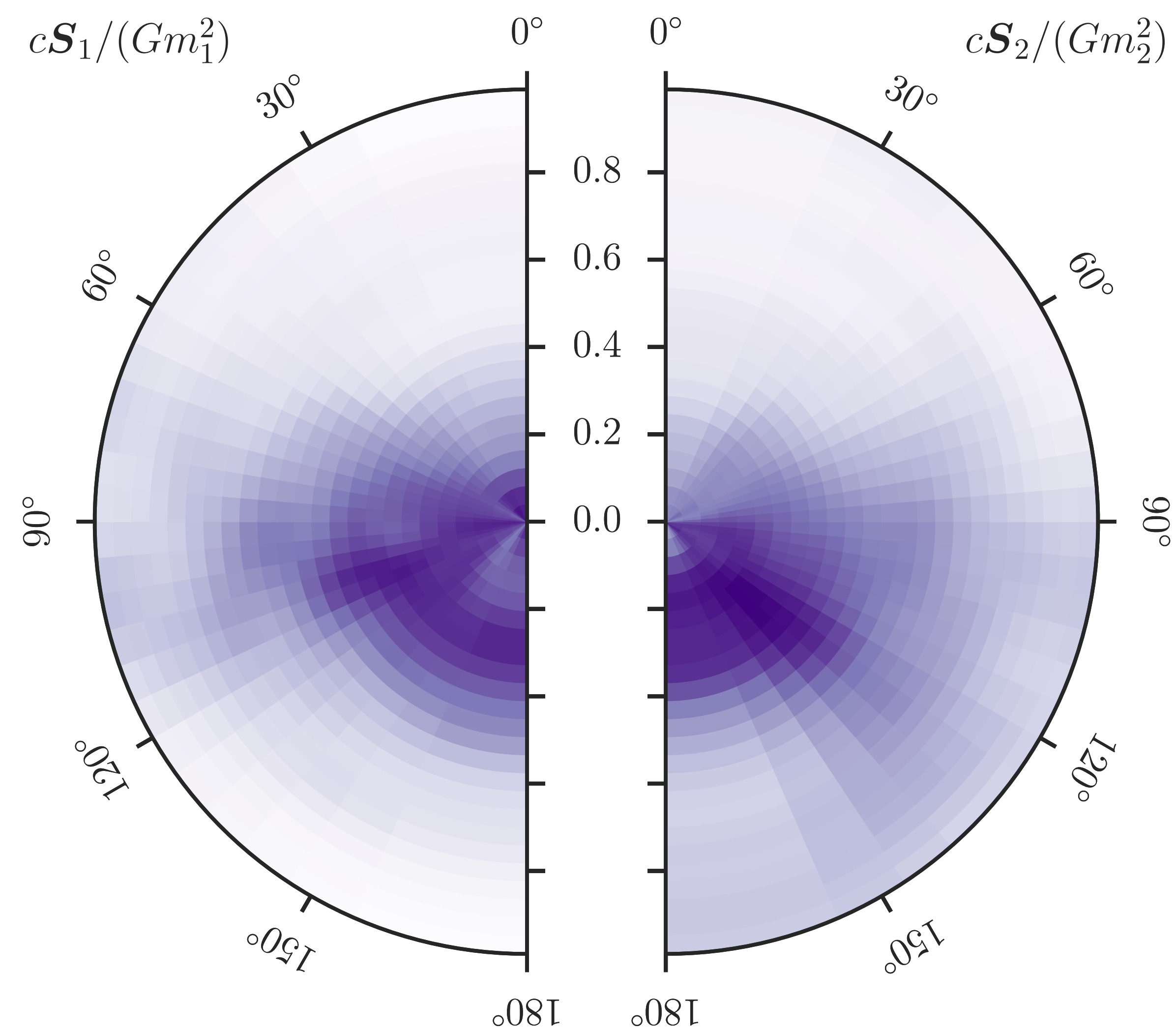
GW150914



Abbott et al, arXiv:1602.03840

GW150914



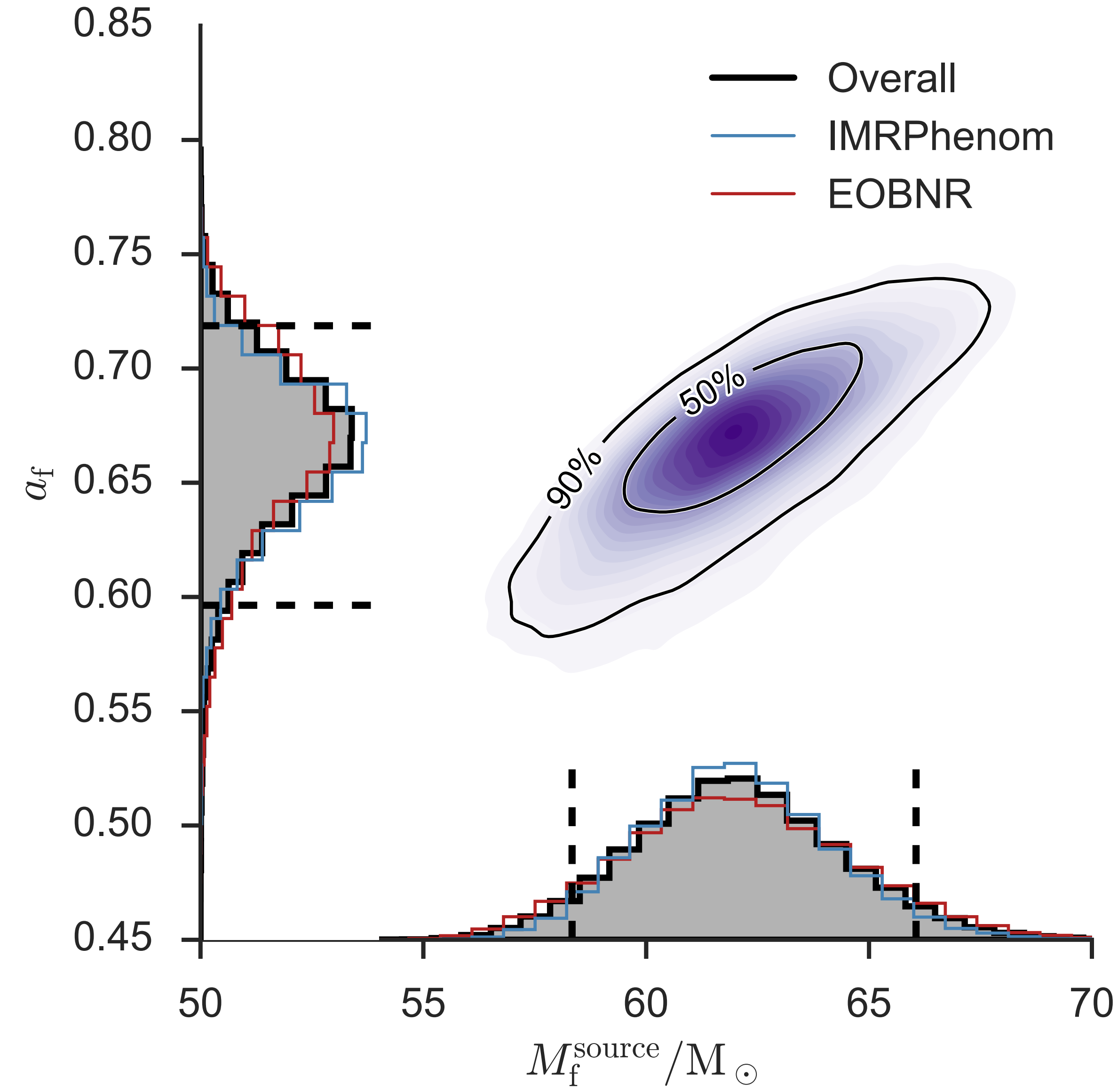


Abbott et al, arXiv:1602.03840

GW150914

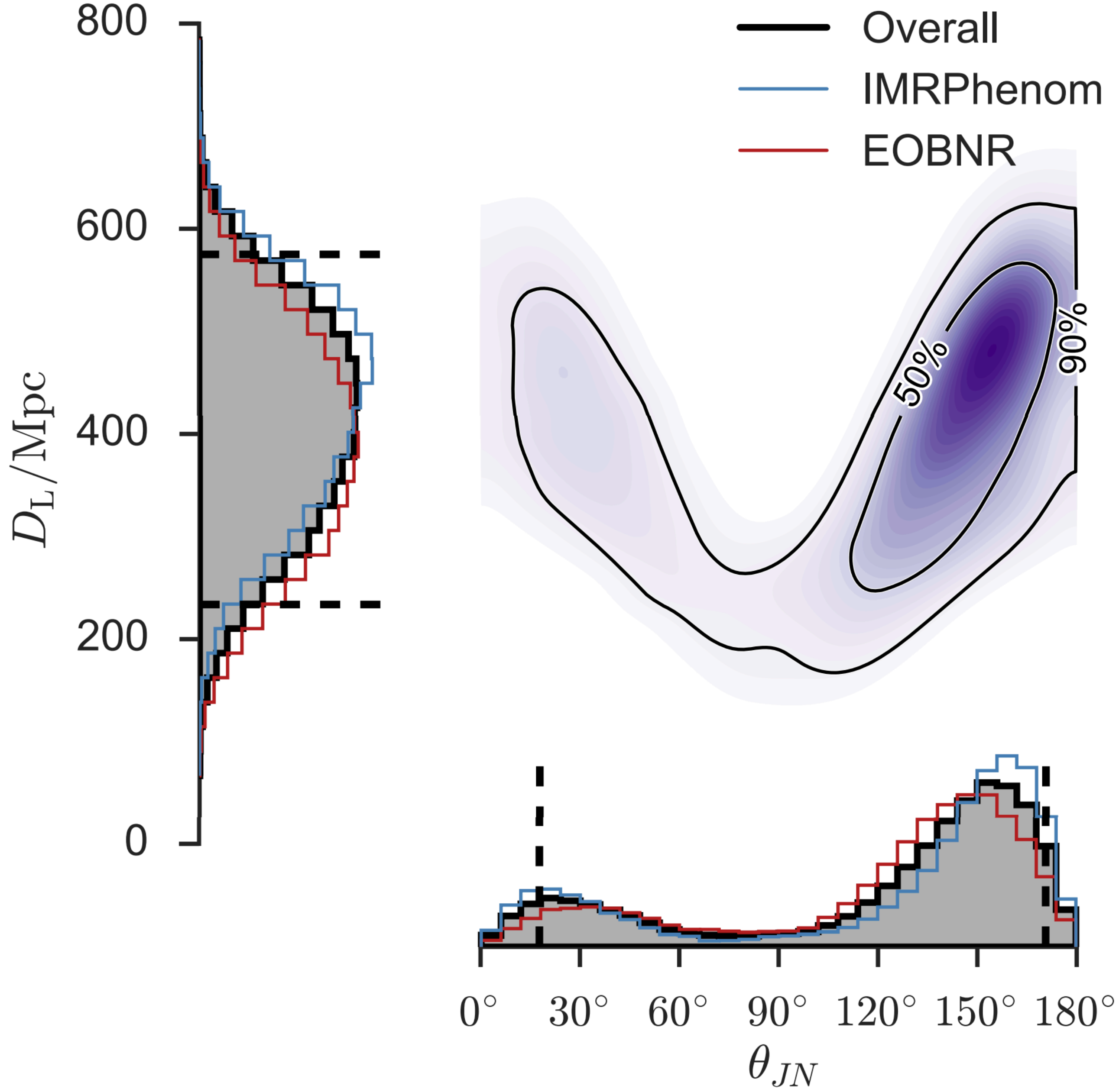
- Accurate measurements require accurate and precise predictions

"Inference is as good as the information you put into it"

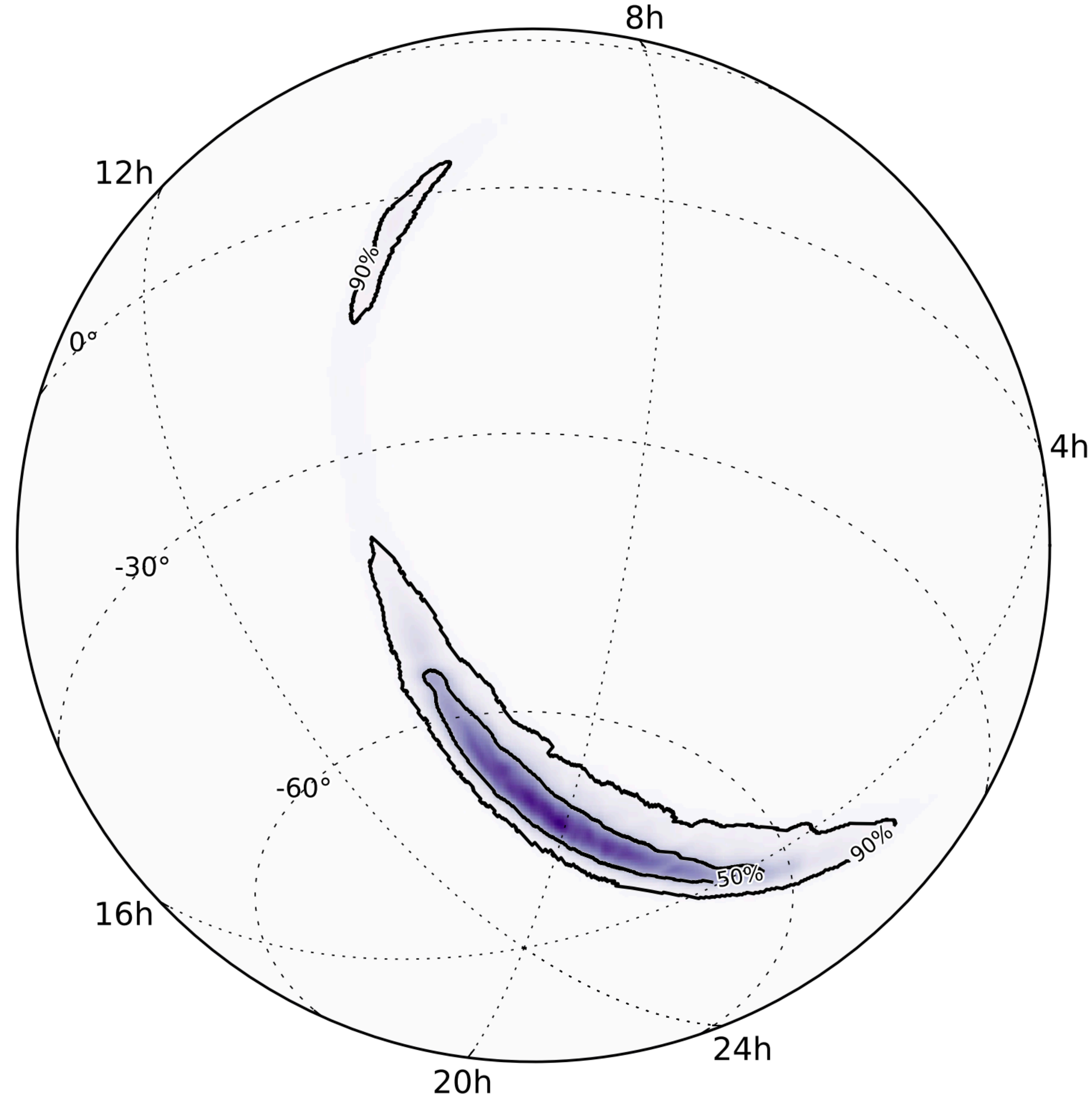


Abbott et al, arXiv:1602.03840

GW150914



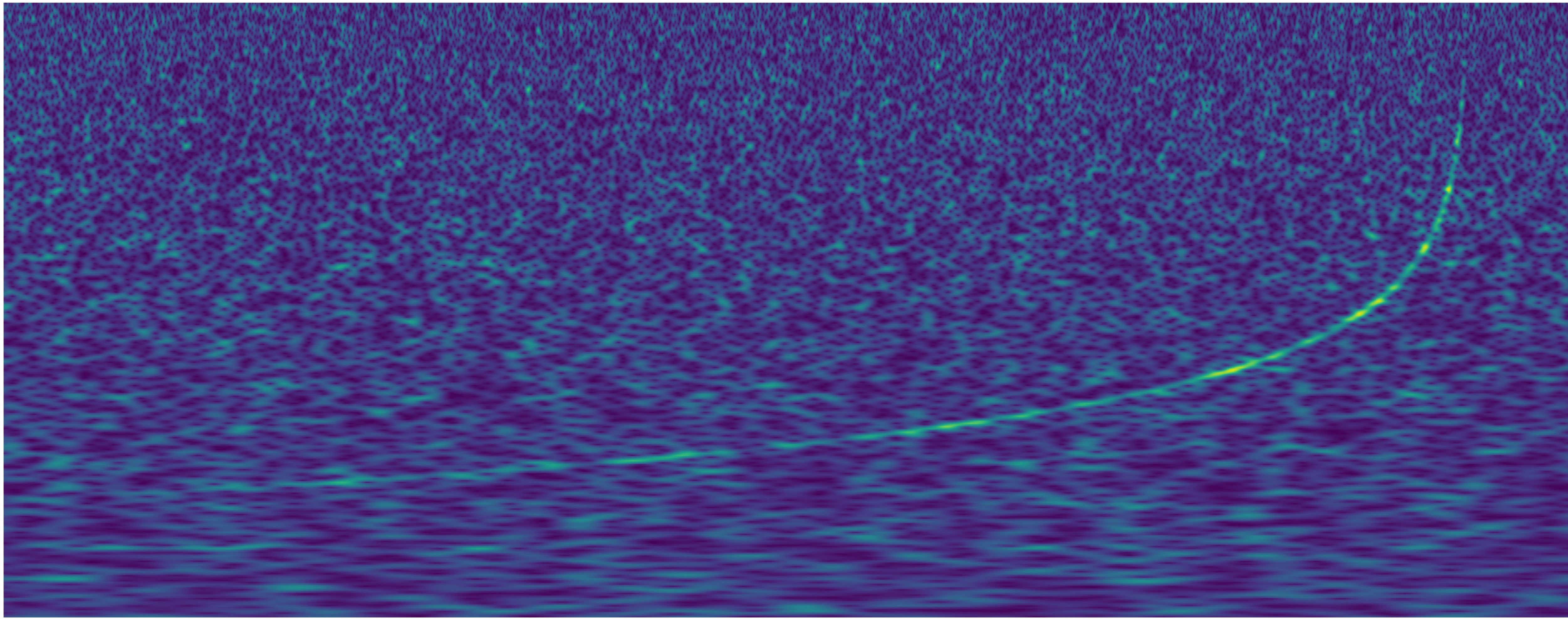
800
600
400
200
0
D_L/Mpc
Overall
IMRPhenom
EOBNR
50%
90%
0°
30°
60°
90°
120°
150°
180°
θ_{JN}



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

GW170817

Frequency

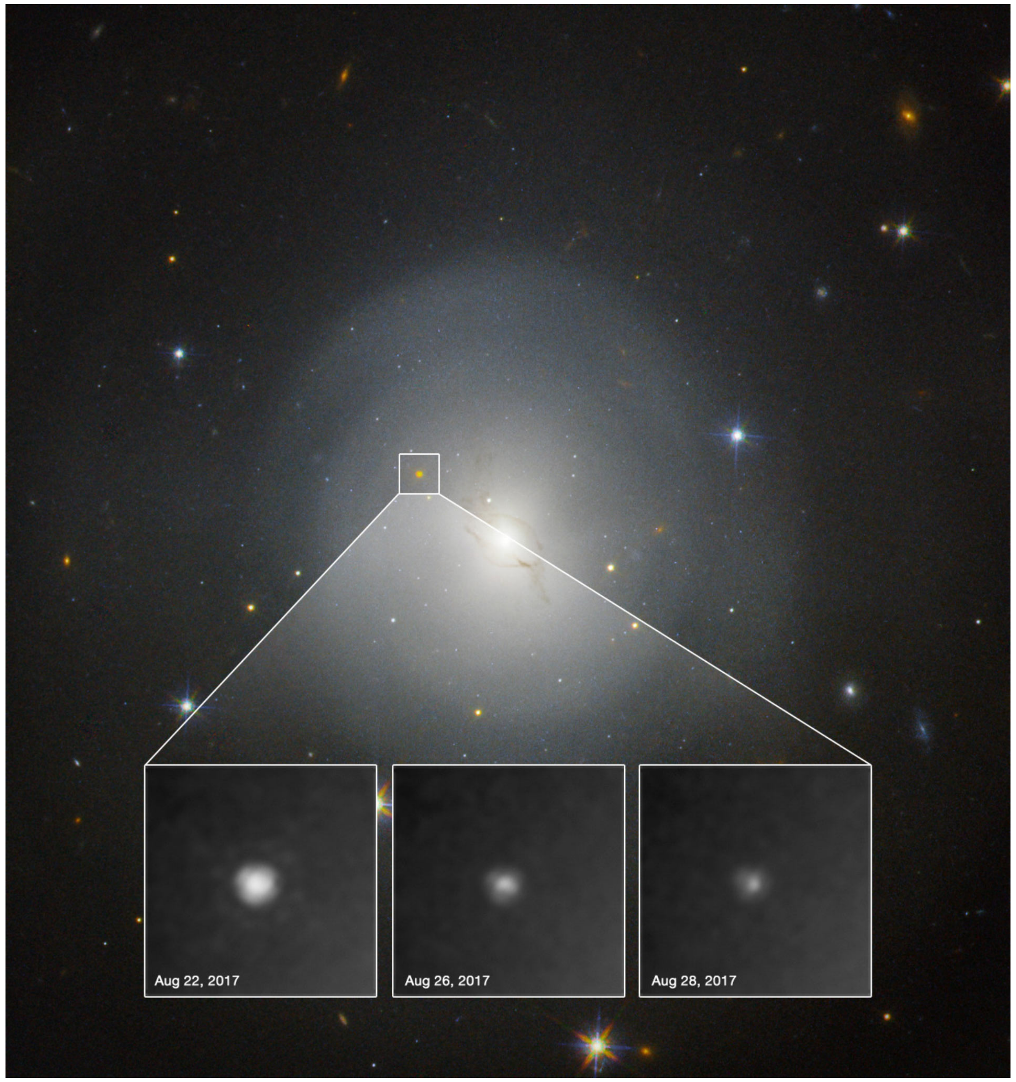


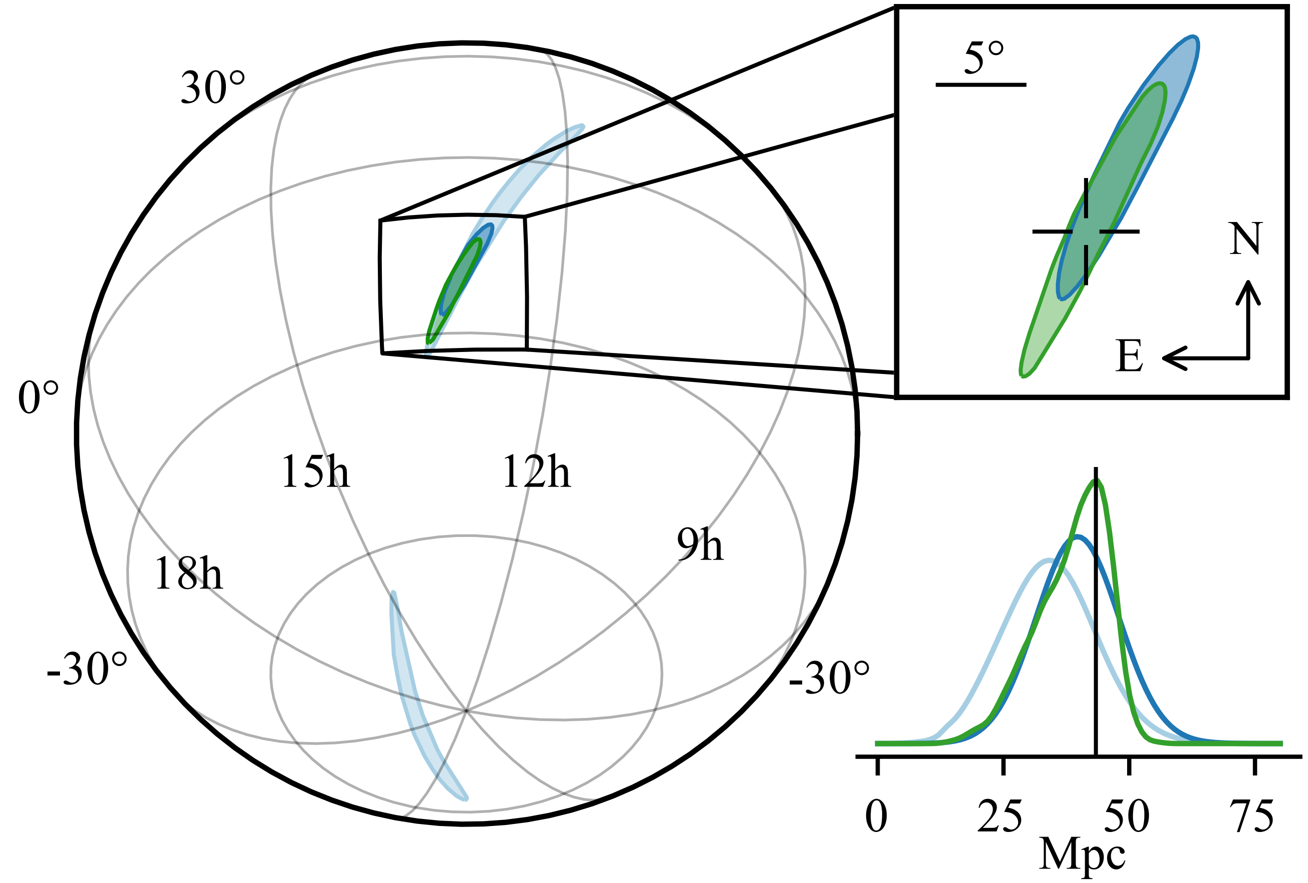
Time

GW170817

	Low-spin priors ($\|\chi\| \leq 0.05$)	High-spin priors ($\|\chi\| \leq 0.89$)
Primary mass m_1	$1.36 - 1.60\ M_\odot$	$1.36 - 2.26\ M_\odot$
Secondary mass m_2	$1.17 - 1.36\ M_\odot$	$0.86 - 1.36\ M_\odot$
Chirp mass $\mathcal{M}$	$1.188^{+0.004}_{-0.002}\ M_\odot$	$1.188^{+0.004}_{-0.002}\ M_\odot$
Mass ratio m_2/m_1	0.7 – 1.0	0.4 – 1.0
Total mass $m_{\rm tot}$	$2.74^{+0.04}_{-0.01}\ M_\odot$	$2.82^{+0.47}_{-0.09}\ M_\odot$
Radiated energy $E_{\rm rad}$	$> 0.025\ M_\odot c^2$	$> 0.025\ M_\odot c^2$
Luminosity distance $D_{\rm L}$	40^{+8}_{-14} Mpc	40^{+8}_{-14} Mpc
Viewing angle Θ	$\leq 55°$	$\leq 56°$
using NGC 4993 location	$\leq 28°$	$\leq 28°$
Combined dimensionless tidal deformability $\tilde{\Lambda}$	≤ 800	≤ 700
Dimensionless tidal deformability $\Lambda(1.4 M_\odot)$	≤ 800	≤ 1400

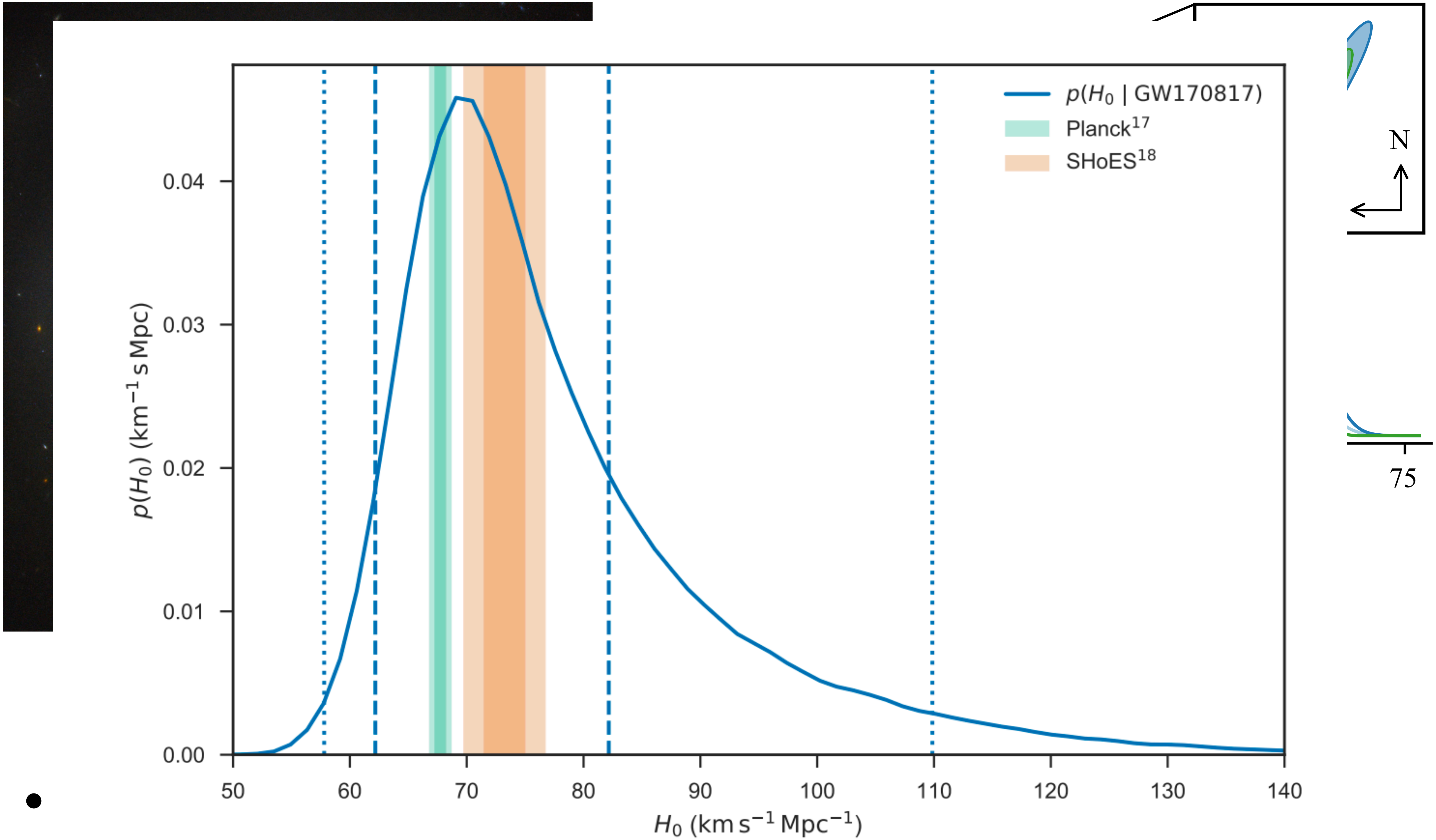
GW170817





- Electromagnetic counterpart!

GW170817



•

LVC, arXiv:1710.05835

Cosmography with GW

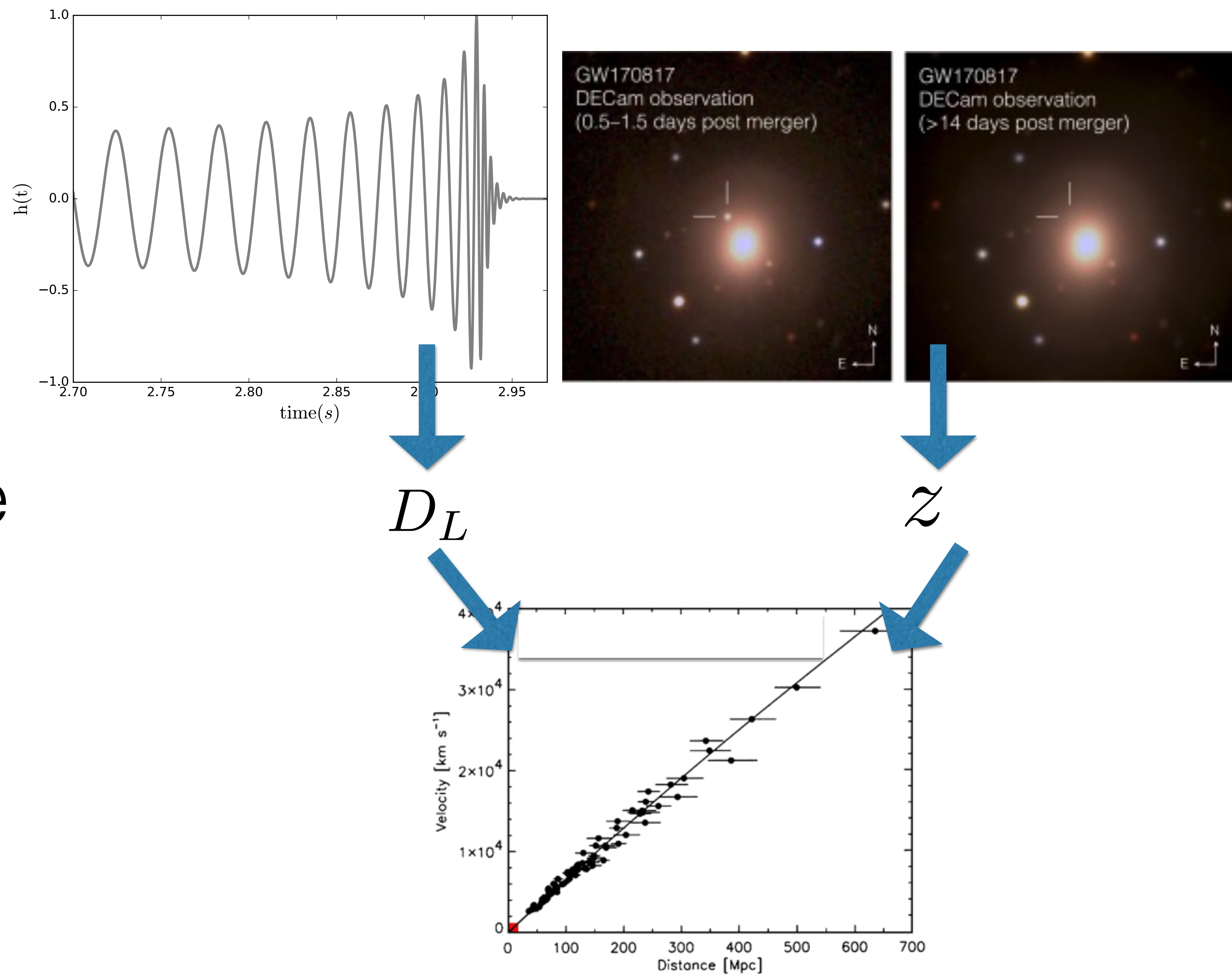
UNIVERSITÀ DI PISA

- GW are self-calibrating

  $h \sim D_L^{-1}$

  - Direct measurement of luminosity distance
  - “Standard sirens”

- In general, no redshift from GWs



Hierarchical models

- How to infer the Hubble constant?

$$p(H_0|DSI) = p(H_0|SI)\frac{p(D|H_0SI)}{p(D|SI)}$$

$$p(D|H_0SI) = \int dD_L dz\, p(D_L\, z|H_0SI)p(D|D_L\, z\, H_0SI)$$

$$= \int dD_L dz\, p(D_L|z\, H_0SI)p(z|H_0SI)p(D|D_L\, z\, H_0SI)$$

$$D_L = \frac{cz}{H_0} \implies p(D_L|z\, H_0\, SI) = \delta(D_L - \frac{cz}{H_0})$$

Del Pozzo, arXiv:1108.1317

Hierarchical models

- Independent measurement of the redshift

$$p(z|H_0\,SI) = p(z|I) = \delta(z - z_g)$$

ignoring uncertainty on redshift from galaxy proper motions, etc…

- Say that

$$p(D|D_L\,z\,H_0\,SI) = p(D|D_L\,SI)$$

$$p(D|D_L\,SI) = e^{-\frac{1}{2}\left(\frac{(D_L-\mu)}{\sigma}\right)^2}$$

- We get

$$p(H_0|DSI) \propto p(H_0|I)e^{-\frac{1}{2}\left[\frac{\left(\frac{cz}{H_0}-\mu\right)}{\sigma}\right]^2}$$

Joint posteriors

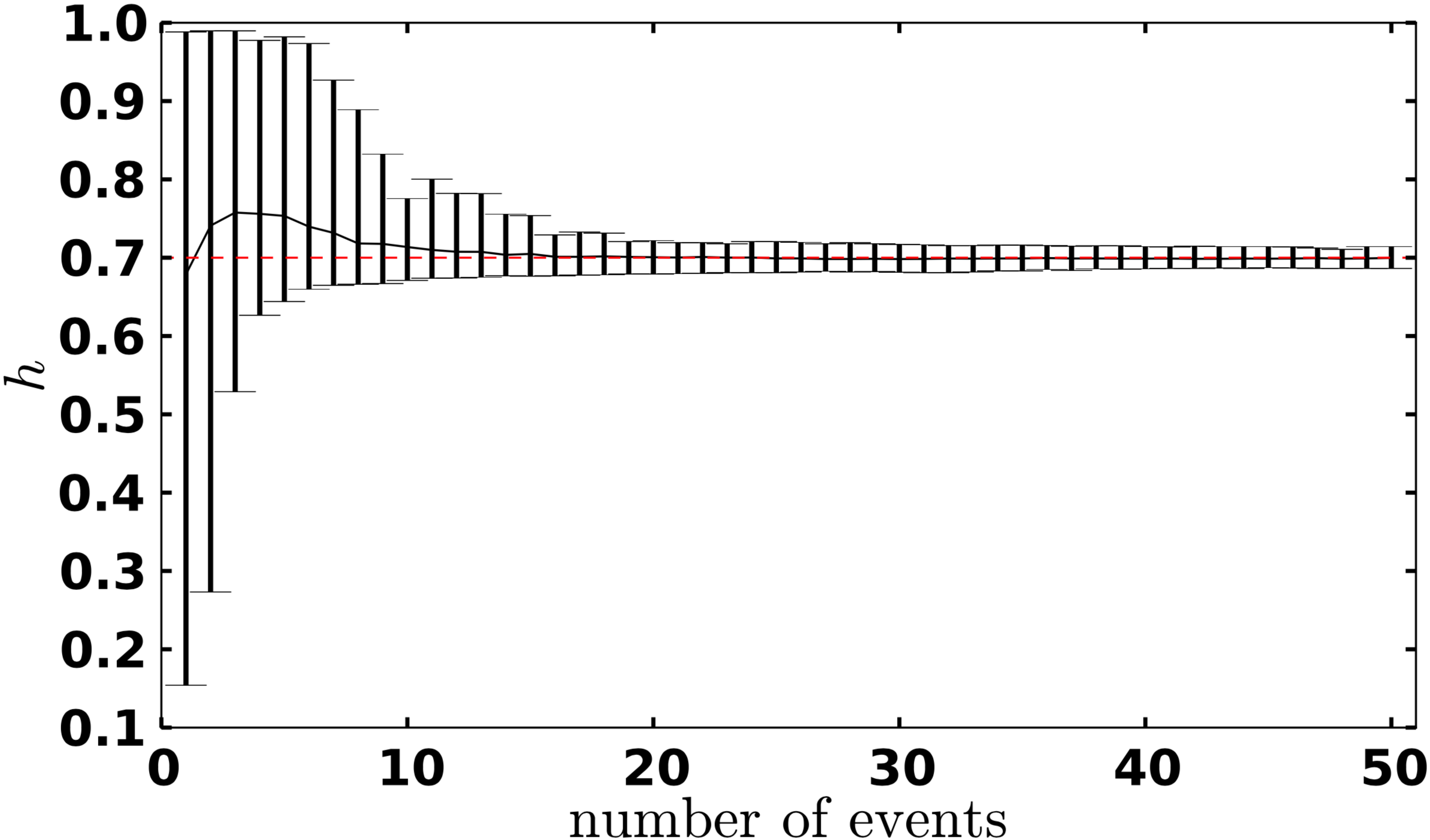
- We have a measurement of a “global” parameter λ
- We measured N events

$$p(\lambda|D_1 \,\ldots\, D_N\, SI) = p(\lambda|SI) \prod_{i=1}^{N} \frac{p(D_i|\lambda\, SI)}{p(D_i|SI)}$$

- Each individual likelihood

$$p(D_i|\lambda\, SI) = \int_{\Theta_i} d\theta_i\, p(\theta_i|\lambda\, SI) p(D_i|\theta_i\, \lambda\, SI)$$

Hubble constant



Del Pozzo, arXiv:1108.1317

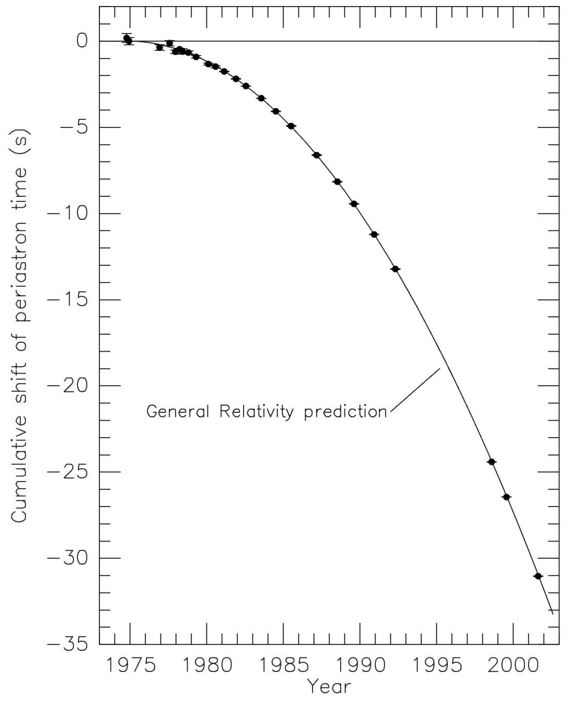
UNIVERSITÀ DI PISA

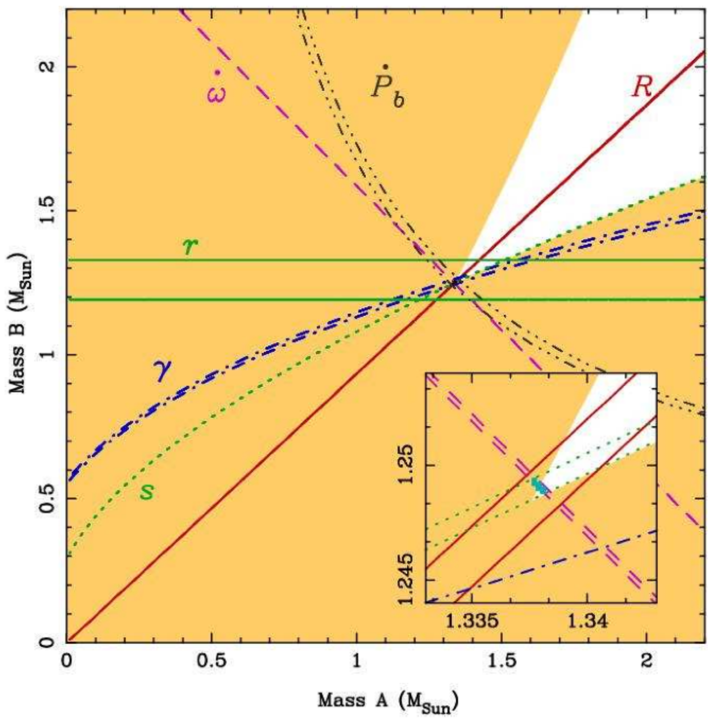
Part 5

Tests of general relativity and black hole nature

Dynamics of space-time

- GR is non renormalisable
  - higher order terms in the action
- Dark matter & dark energy
  - signature of modified gravity?
- GR is extremely well tested in between these regimes (Will, arXiv:1403.7377, Psaltis, arXiv:0806.1531)





Weisberg & Taylor, arXiv:0407149
Kramer+,arXiv:0609417

Gravitational strong-field

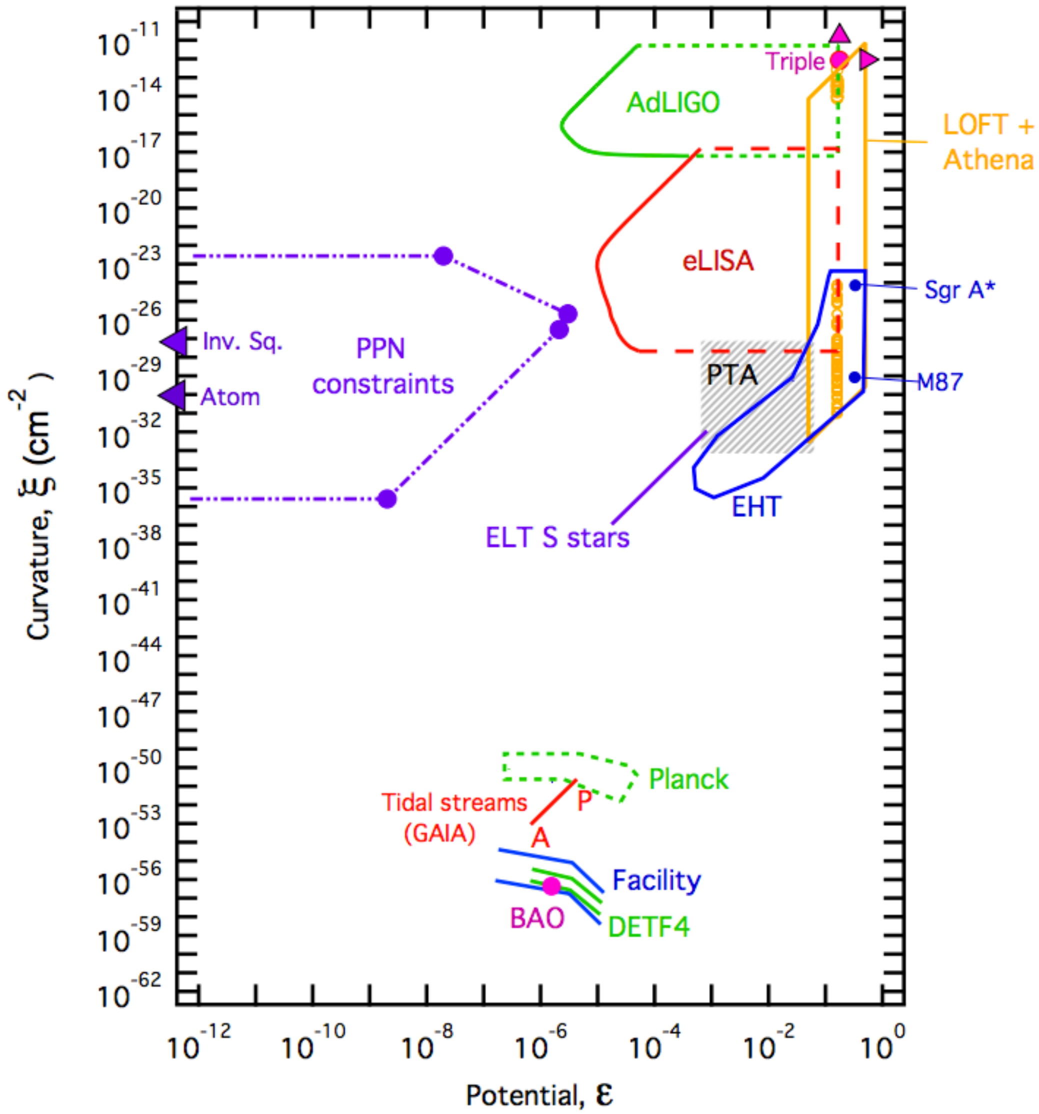
- Field strength

$$\epsilon = \frac{GM}{c^2 R}$$

- Curvature (Kretschmann scalar)

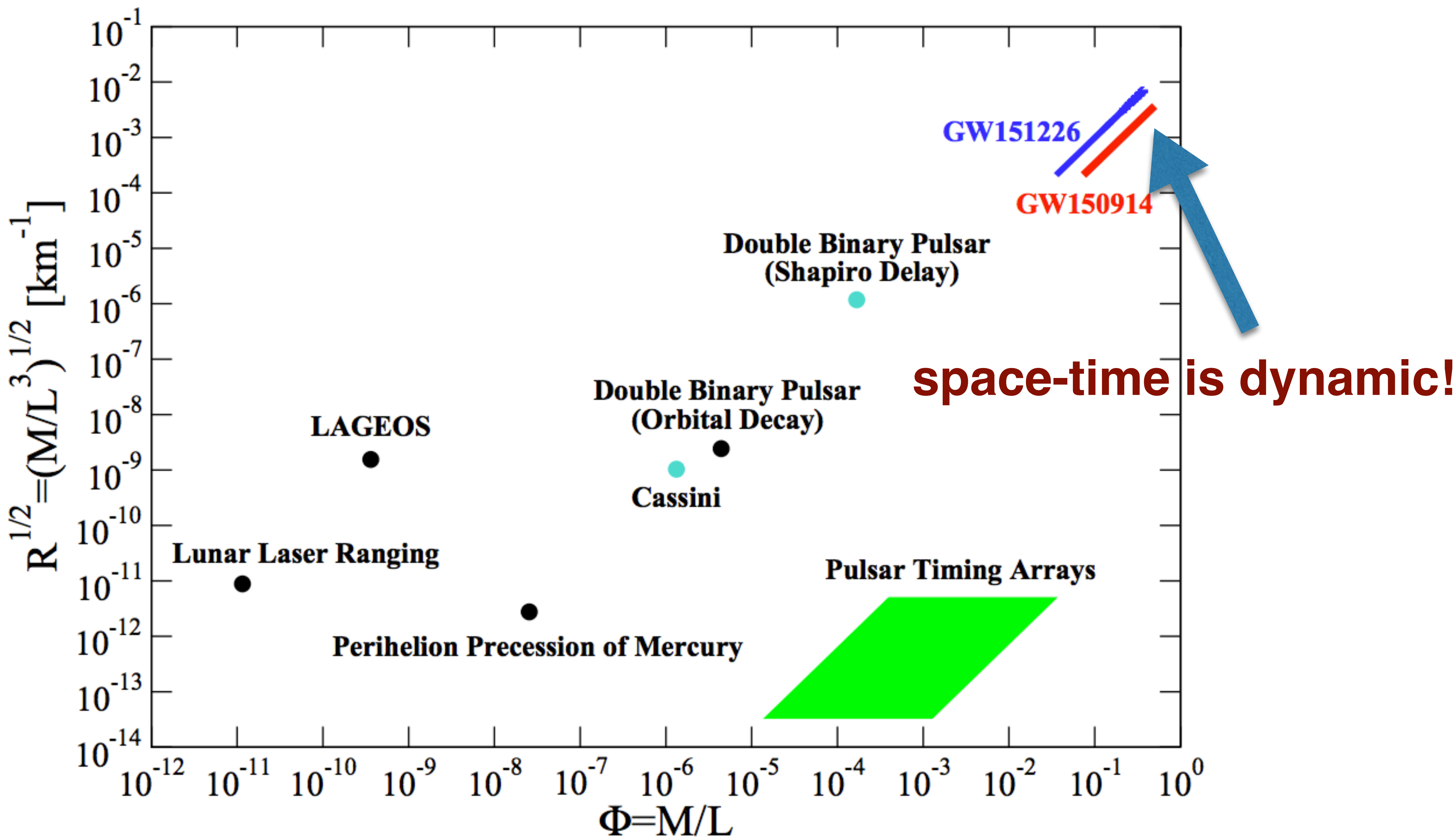
$$\xi = (R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta})^{1/2}$$

- Gravitational waves from binary black holes are the optimal probes



Baker et al,arXiv:1412.3455

Gravitational strong-field

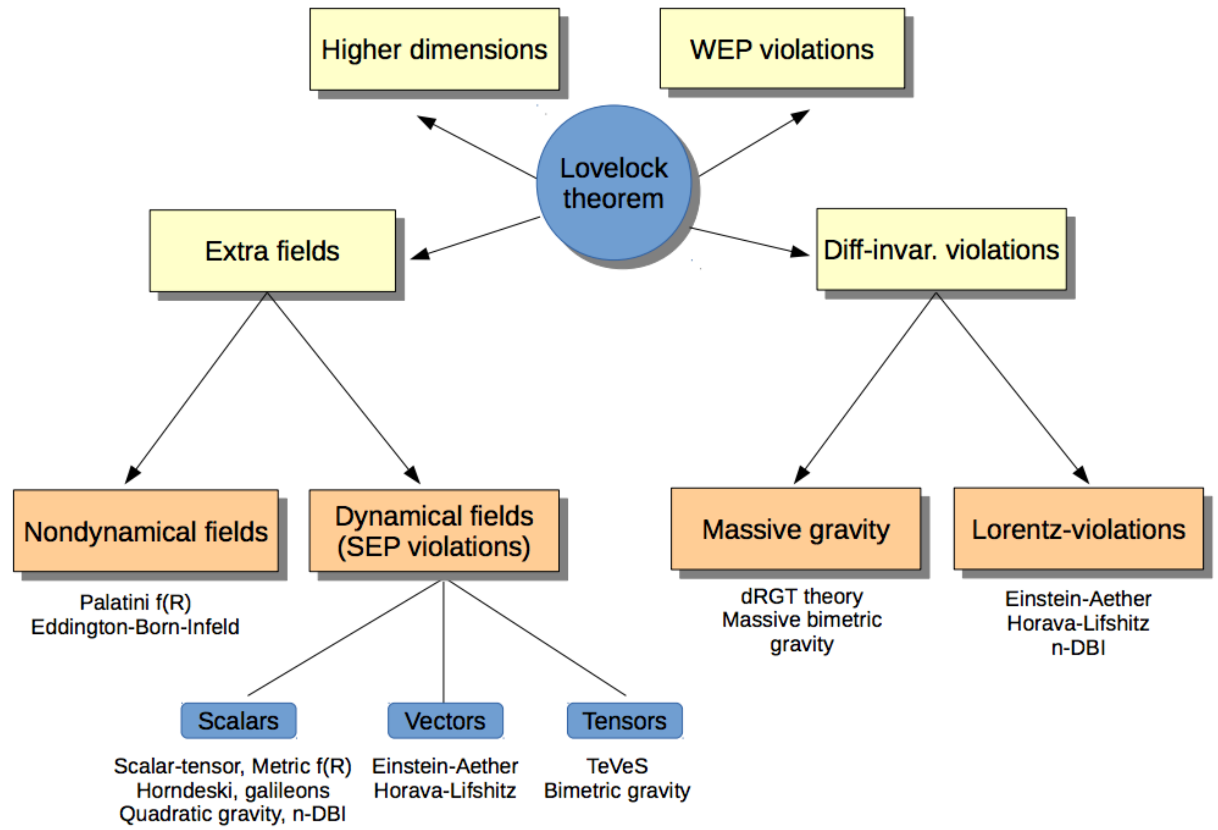


Yunes et al,arXiv:1603.08955

Extensions of GR

- Alternative theories
  - Introduce extra degrees of freedom:
    - additional fields
    - higher-curvature terms
  - Challenge GR assumptions:
    - Lorentz invariance
    - Equivalence principle
- Need tests in the strong-field

Lovelock theorem: In 4D, the only divergence free symmetric rank-2 tensor constructed only by the metric and its derivatives up to 2nd order and preserving diffeomorphism invariance is the Einstein tensor plus a constant.



Berti et al, arXiv:1501.07274

non-GR effects on the waveform

- Alternative theories of gravity modify the waveform
  - change the φ coefficients by introducing additional parameters
    - e.g. "massive gravity"
  - add extra orders not present in the GR waveform
    - e.g. Brans-Dicke
- Non-BHs show different merger and ringdown spectra

$$h(f) = A(f)e^{i\Phi(f)}$$

$$\Phi(f) = \sum_{k=1}^{7}(\varphi_k + \varphi_k^l \log(f))f^{(5-k)/3} + \sum_{i\neq k}\varphi_i f^i$$

$$\varphi_j \equiv \varphi_j(m_1, m_2, \vec{s}_1, \vec{s}_2)\ \forall\, j = k, i$$

$$\Phi_{MG}(v) = \Phi_{GR}(v) - \frac{\pi^2 DM}{\lambda_g^2(1+z)}v^{-1}$$

$$\Phi_{\mathrm{BD}}(v) = \Phi_{\mathrm{GR}}(v) - \frac{5S^2}{84\omega_{BD}}v^{-2}$$

Strategies

- Two strategies to test GR

  1. Self-consistency tests: perturb around GR and check for evidence of inconsistencies

  2. Targeted tests: assume an alternative theory of gravity and constrain its parameters

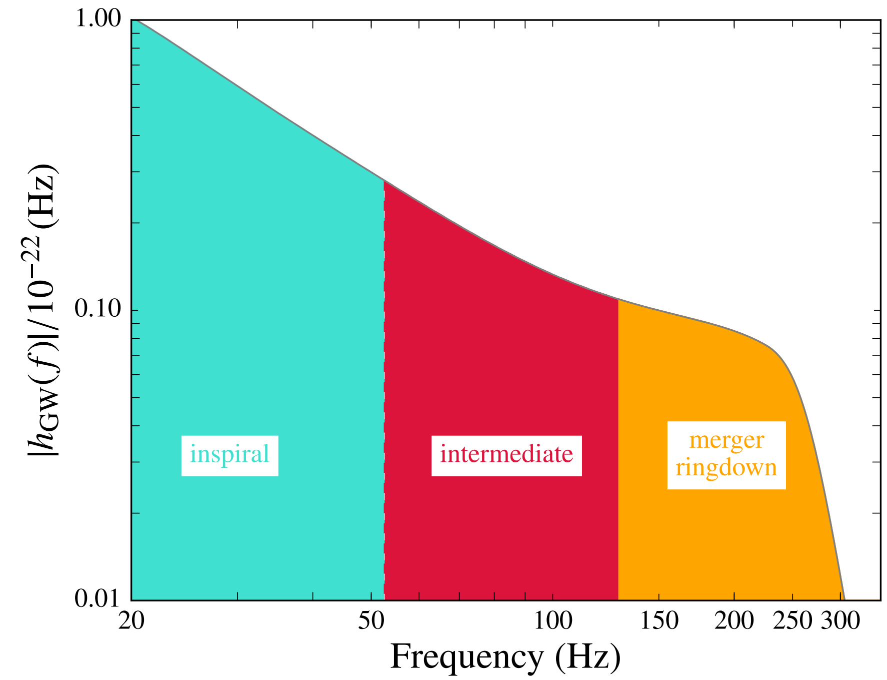
Parametrised tests of GR

UNIVERSITÀ DI PISA

- Waveform models are described by post-Newtonian and phenomenological coefficients
- Allow for fractional changes with respect to the GR value

$$\hat{\varphi}_j \equiv \varphi^{\rm GR}(1+\delta\hat{\varphi}_j)$$

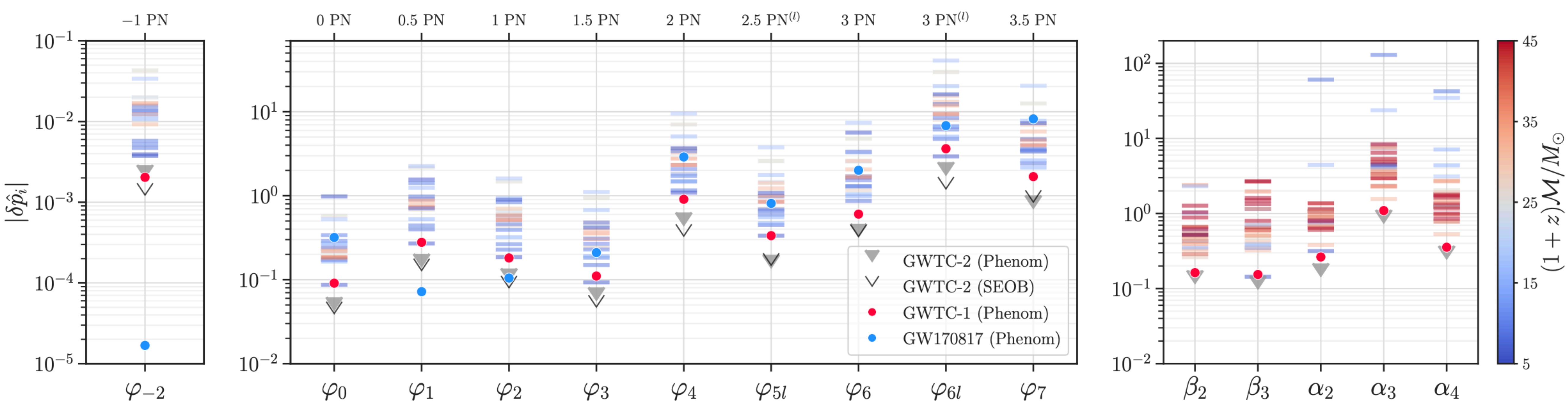
- Obtain constraints on "generic" deviations from GR (Li, Del Pozzo et al 2012, Agathos, Del Pozzo et al 2013)



waveform regime	parameter	f–dependence
early-inspiral regime	$\delta\hat{\varphi}_0$	$f^{-5/3}$
	$\delta\hat{\varphi}_1$	$f^{-4/3}$
	$\delta\hat{\varphi}_2$	f^{-1}
	$\delta\hat{\varphi}_3$	$f^{-2/3}$
	$\delta\hat{\varphi}_4$	$f^{-1/3}$
	$\delta\hat{\varphi}_{5l}$	$\log(f)$
	$\delta\hat{\varphi}_6$	$f^{1/3}$
	$\delta\hat{\varphi}_{6l}$	$f^{1/3}\log(f)$
	$\delta\hat{\varphi}_7$	$f^{2/3}$
intermediate regime	$\delta\hat{\beta}_2$	$\log f$
	$\delta\hat{\beta}_3$	f^{-3}
merger–ringdown regime	$\delta\hat{\alpha}_2$	f^{-1}
	$\delta\hat{\alpha}_3$	$f^{3/4}$
	$\delta\hat{\alpha}_4$	$\tan^{-1}(af+b)$

post-Newtonian series

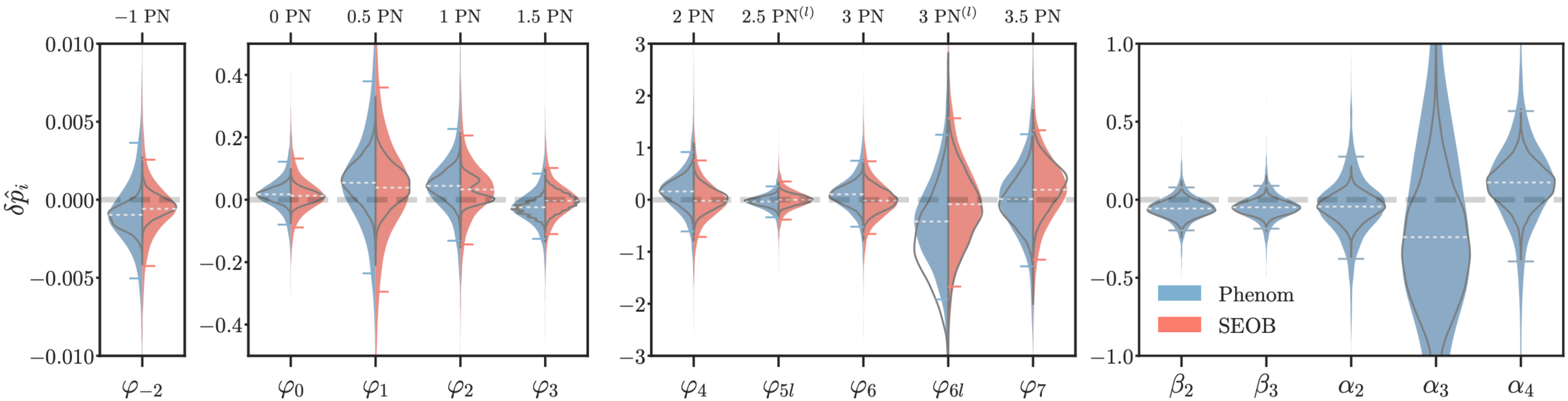
- Dynamical constraints on post-Newtonian series
- Constraints on non-linear dynamics of space-time during the inspiral



LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

UNIVERSITÀ DI PISA

Parametrised tests of GR



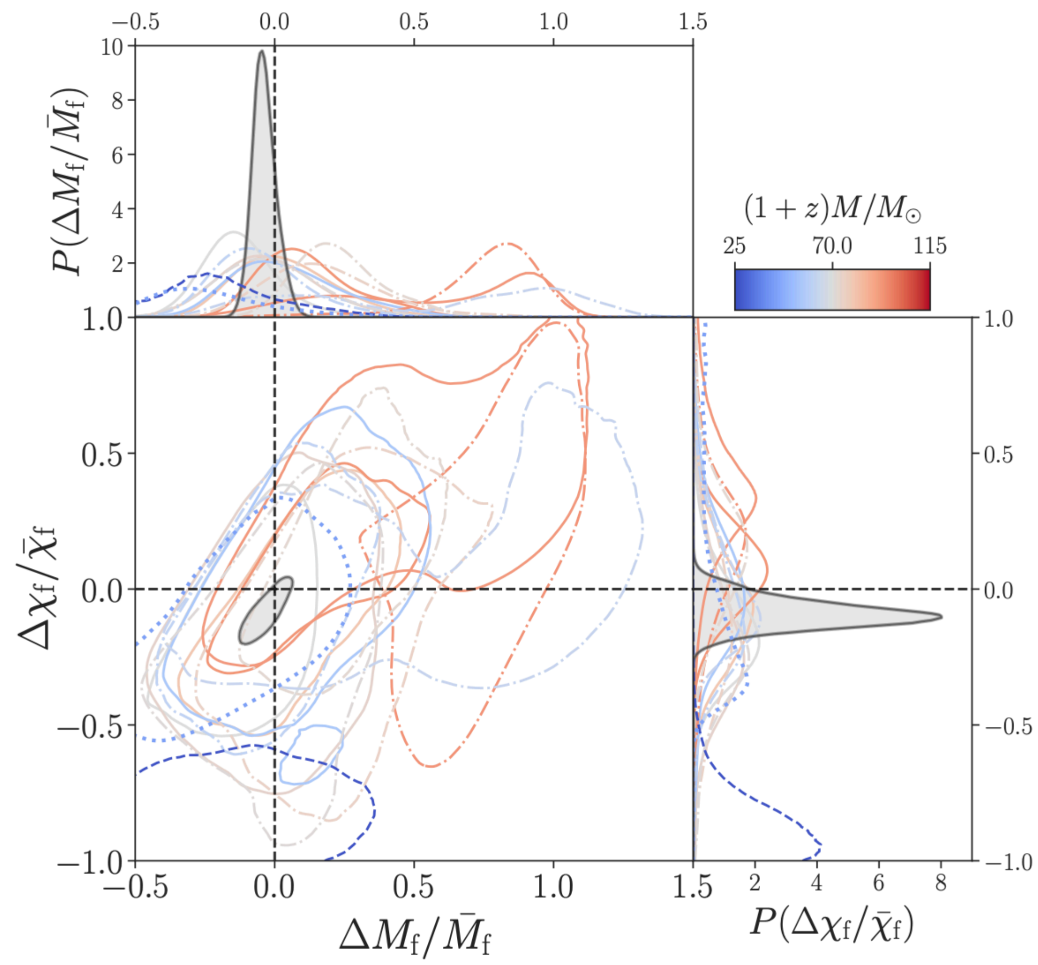
Energy/Frequency

LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

Reconstructed waveform consistency

$$\begin{cases} \Delta M_f / M_f = 0 \\ \Delta a_f / a_f = 0 \end{cases} \iff GR$$

- If GR is verified, the recovered GR waveform must be self-consistent
- Numerical solution provide predictions for spin and mass of remnant starting from the parents ones (e.g. Healy+, 1406.7295)
- Verify self-consistency by comparing final mass and spin predicted from the “inspiral” with the ones inferred from the “post-inspiral” (Ghosh+,1602.02453)



LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

Propagation tests: massive gravity

- Families of alternative theories modify the propagation of GW
- Massive gravity (e.g. Will, arXiv:9709011)

$$E^2 = p^2 v_g^2 + m_g^2 c^4$$

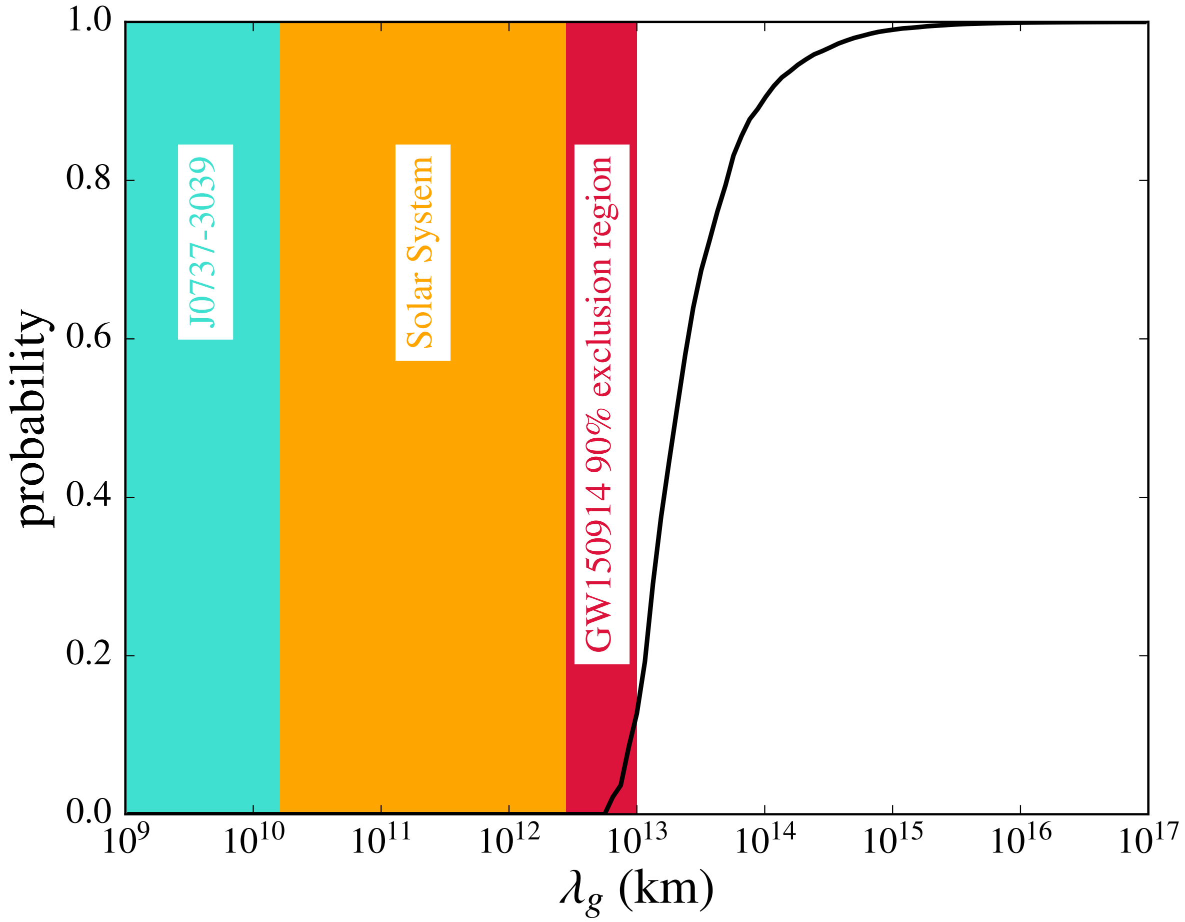
$$v_g^2/c^2 \simeq 1 - \frac{h^2 c^2}{\lambda_g^2 E^2} \qquad \lambda_g = \frac{h}{m_g c}$$

- GW phase affected

$$\Delta\Phi = -\frac{\pi^2 DM}{\lambda_g^2 (1+z)}$$

- GW constrains gravitons Compton wavelength

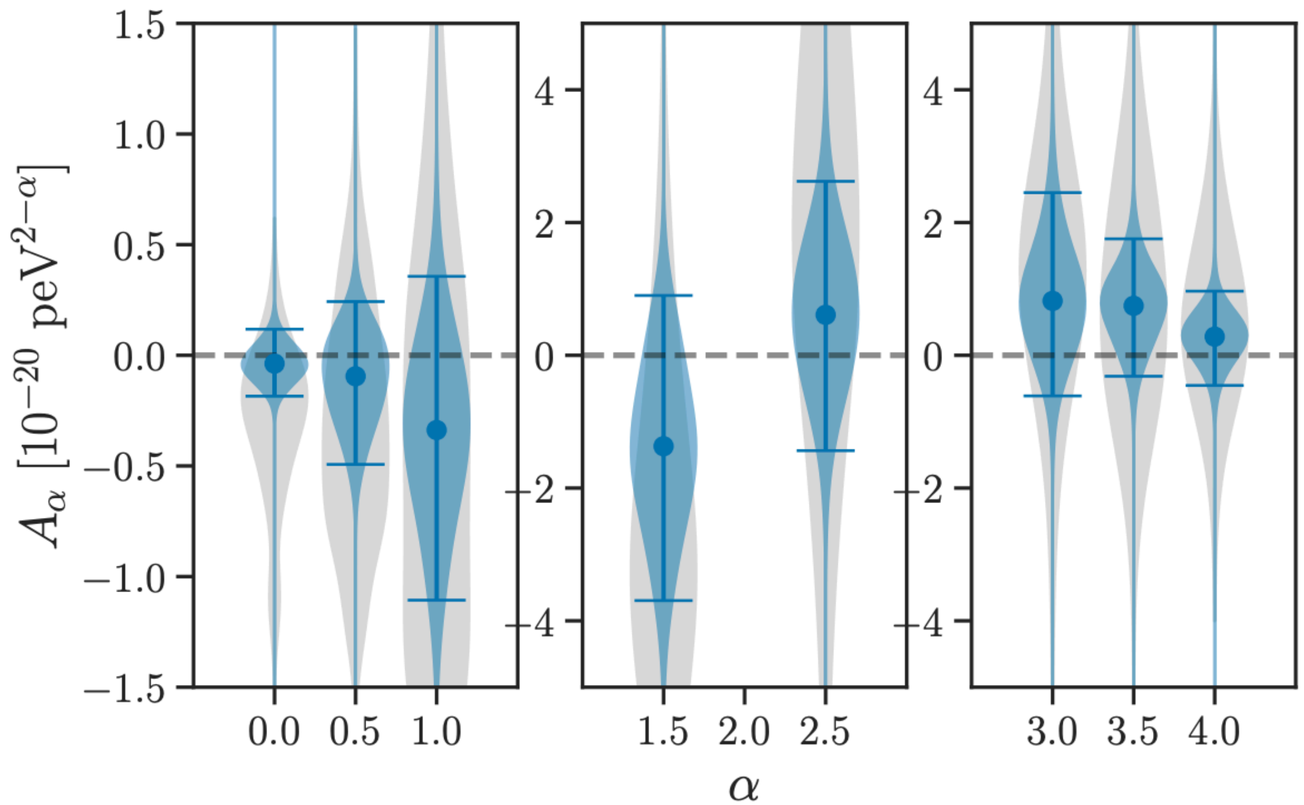
LVC, arXiv:1602.03841



Tests of Lorentz Invariance Violations

- Further generalised (e.g. Mirshekari et al, arXiv:1110.2720)

$$E^2 = p^2c^2 + Ap^\alpha c^\alpha \quad \alpha \geq 0$$
$$v_g/c = 1 + (\alpha - 1)AE^{\alpha-2}/2$$



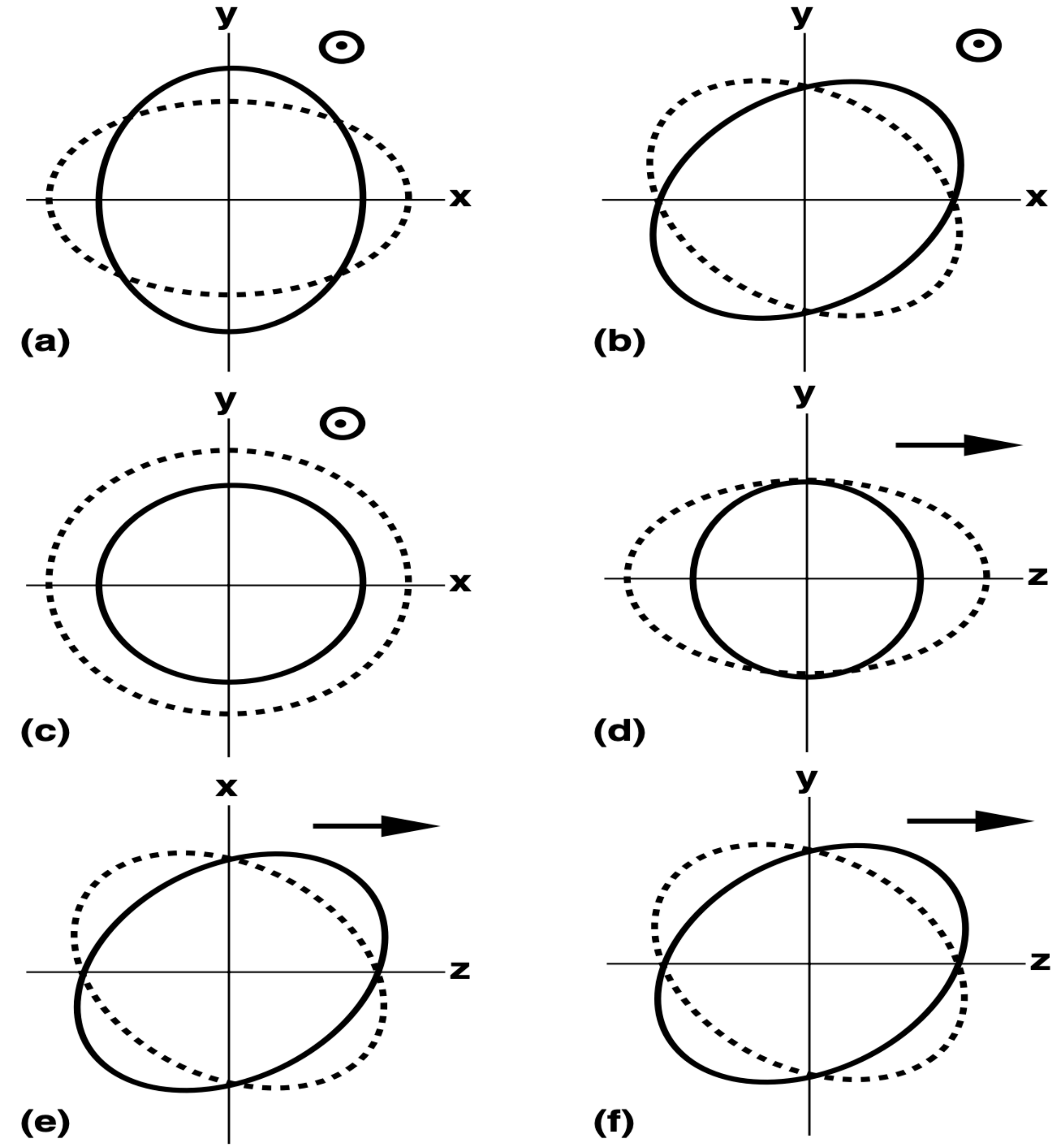
- first bounds derived from GW

- first tests of superluminal propagation in the gravitational sector

LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

Gravitational wave polarisation states

- Gravitational waves in general relativity are transverse, tensorial waves
- Extensions to general relativity predict up to six polarisation states
  - Two transverse tensor states
  - Two longitudinal vector states
  - Two scalar states, one longitudinal and one "breathing"

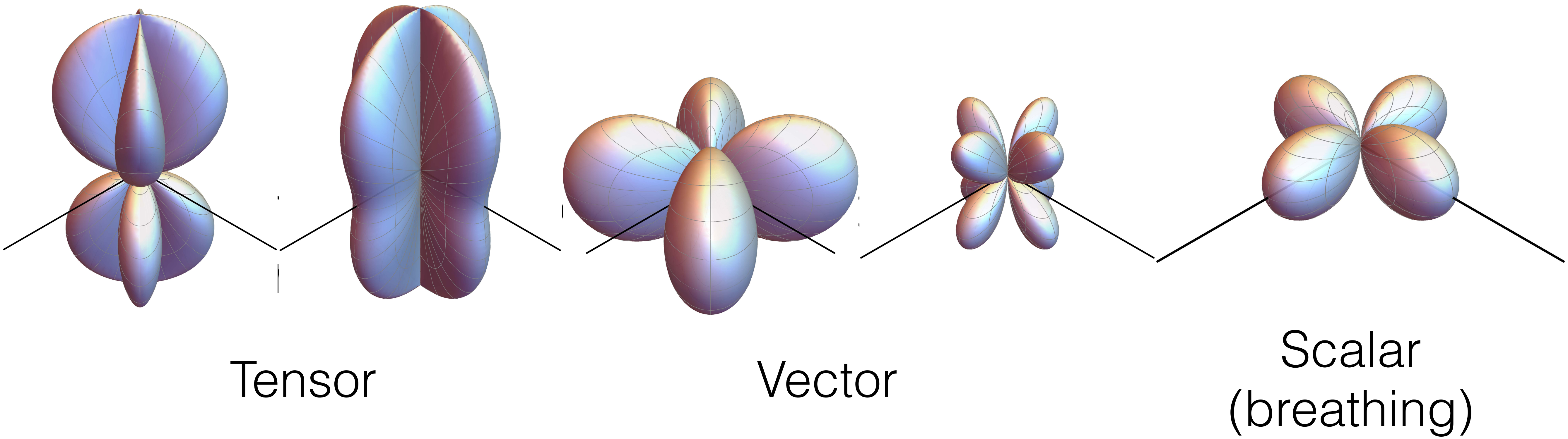


Capozziello & De Laurentis, 2012

Detector response to polarisation states

$$h = \sum_{k=1}^{6} F_k h_k$$

Antenna response functions F_k

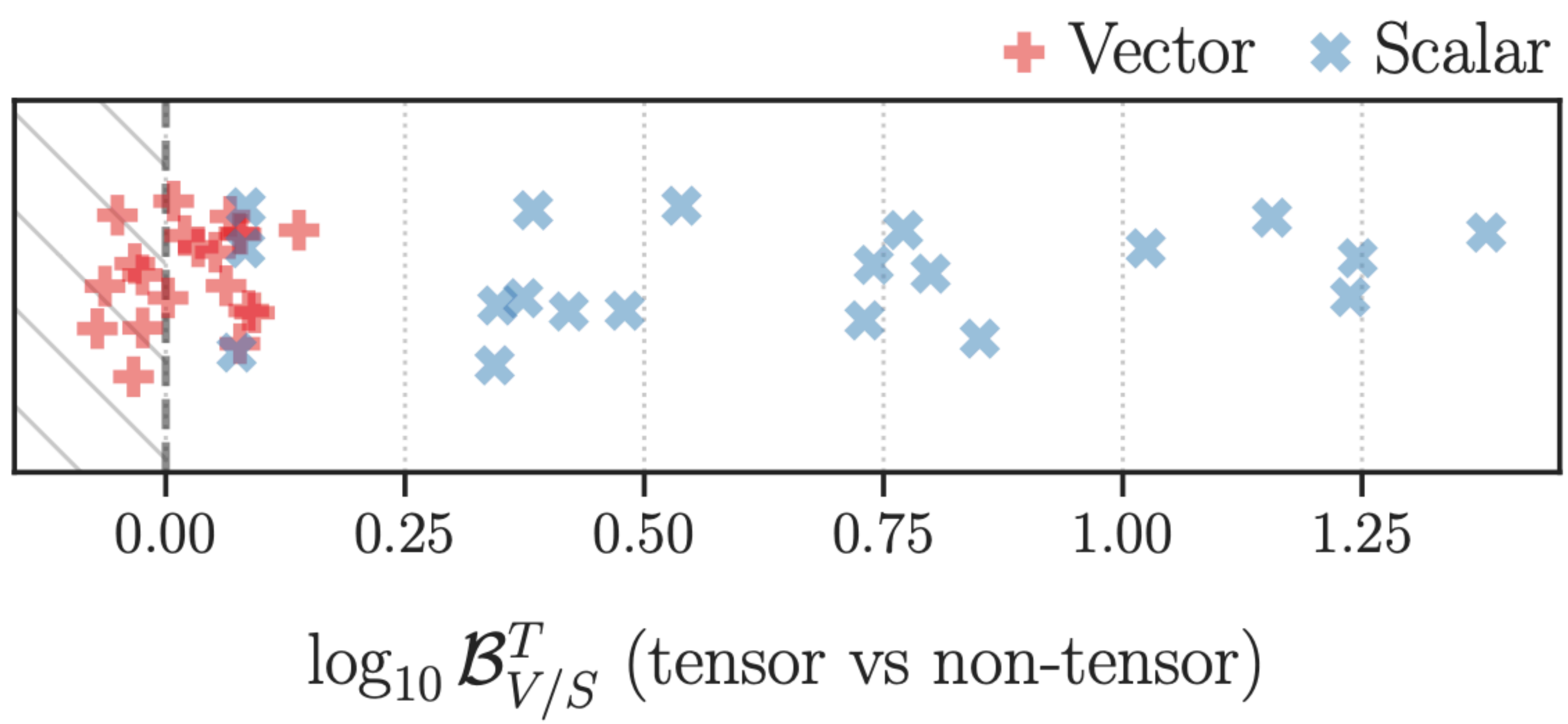


- In principle detectable with more than one detector

Courtesy of Max Isi

Two detectors sensitivity to polarisation states

- The two LIGO detectors cannot not discriminate among different polarisation states
- With Virgo detector and/or an electromagnetic counterpart we can



LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

On the nature of black holes

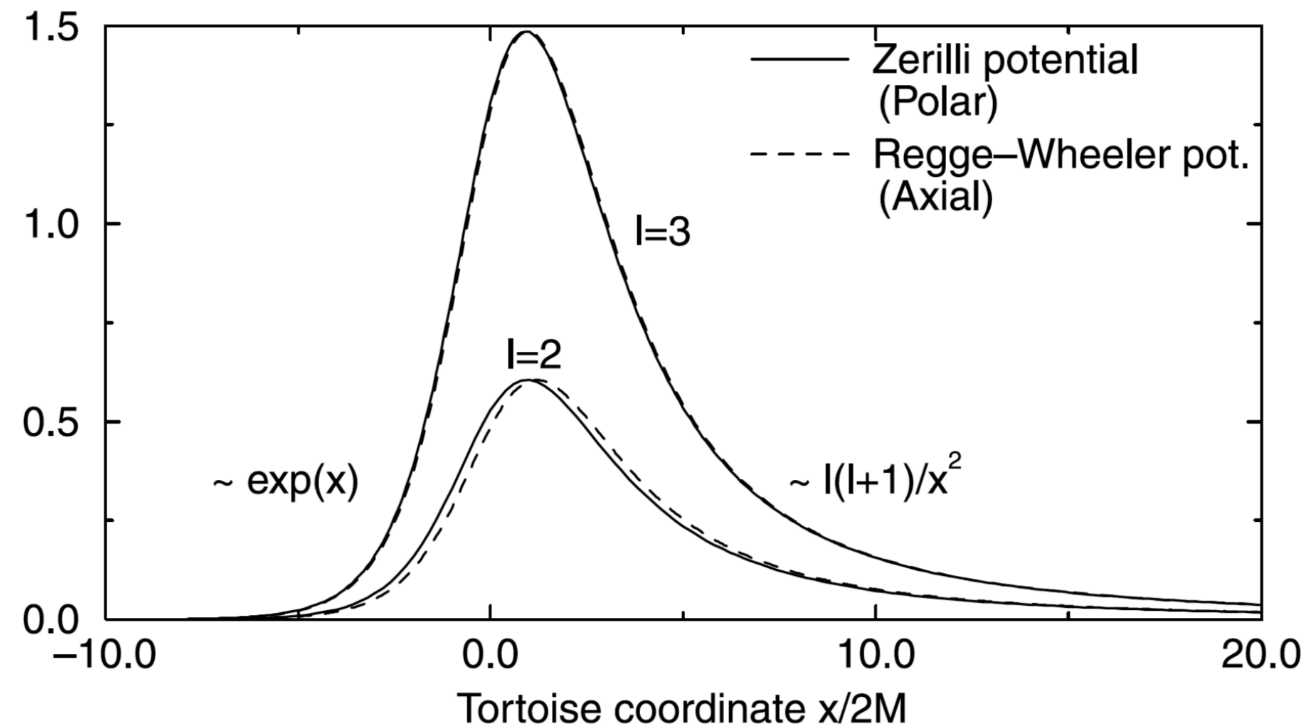
BH perturbation theory

- Study linear perturbation around Schwarzschild and Kerr metrics

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + V_{lm}(x)\right)\psi_{lm}(x,t) = 0$$

- Perturbations obey Schrodinger-like equation (Regge & Wheeler 56, Zerilli 70, Teukolski 72)

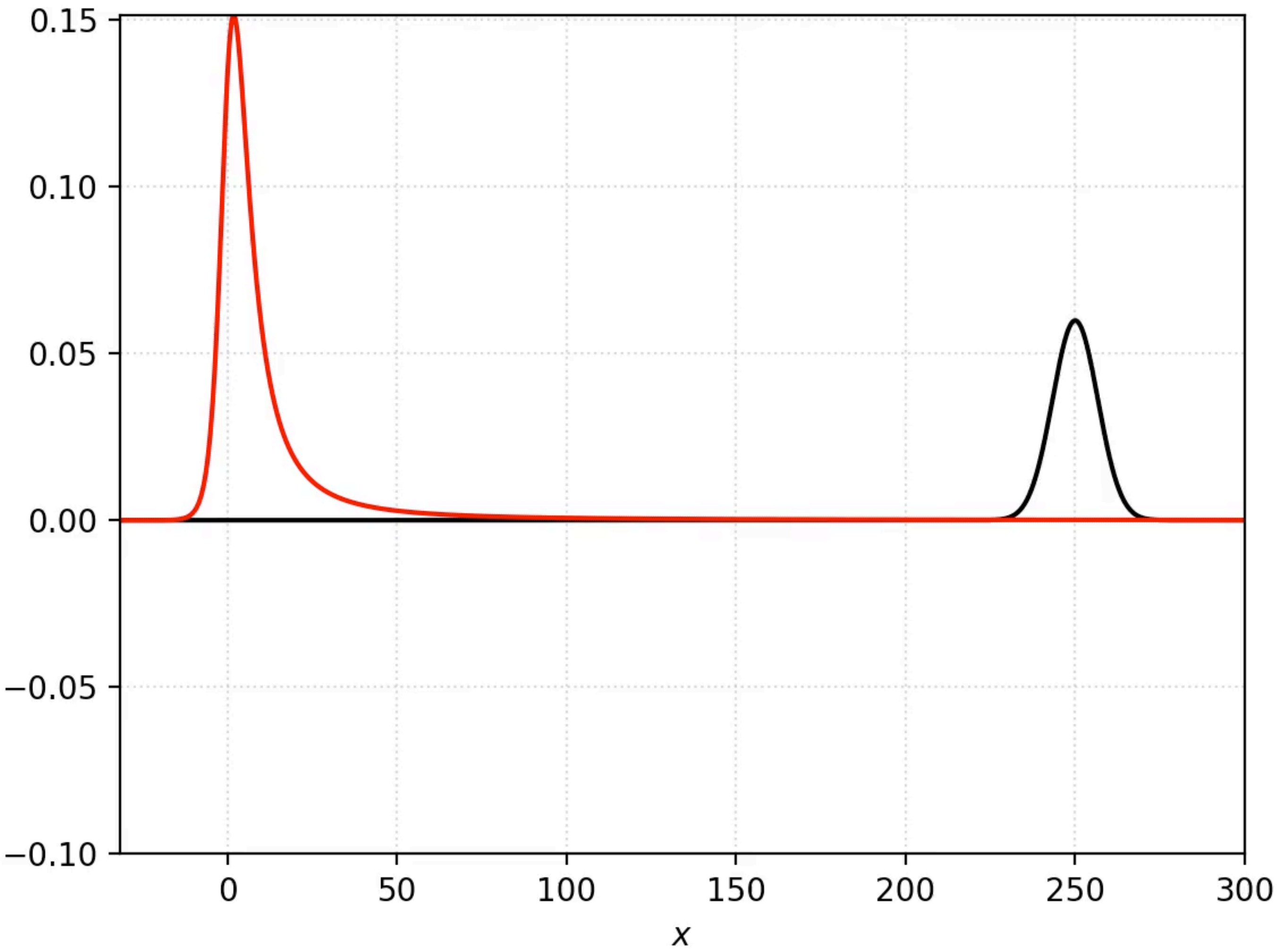
- Potential barrier around a BH



Nollert, 1999

Quasi-normal modes

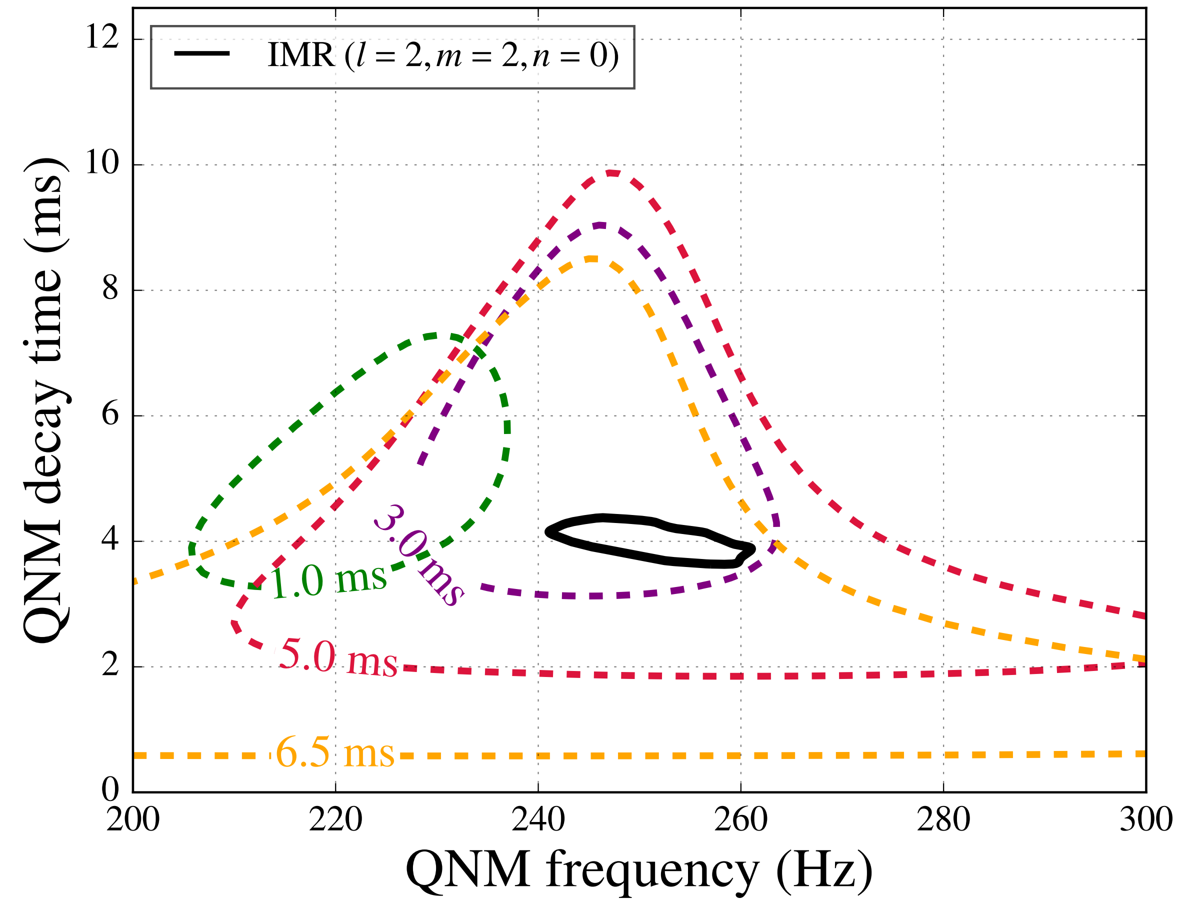
- BH responds to perturbations by "ringing" (Vishveshawara 70, Press 71, Ruffini et al, 72, Chandrasekhar 75)
  - Quasi-normal modes excited by light-ring crossing (Goebel 72)



- Waveform $$h(t) = \sum_{nlm} A_{nlm} e^{-\frac{t-t_0}{\tau_{nlm}}} \cos(\omega_{nlm}(t-t_0) + \varphi_{nlm})$$

The nature of the final object

- Central frequencies ω_{nlm} and decay times τ_{nlm} are functions of BH mass and spin only (manifestation of the BH uniqueness hypothesis, Berti et al, arXiv:0512160)

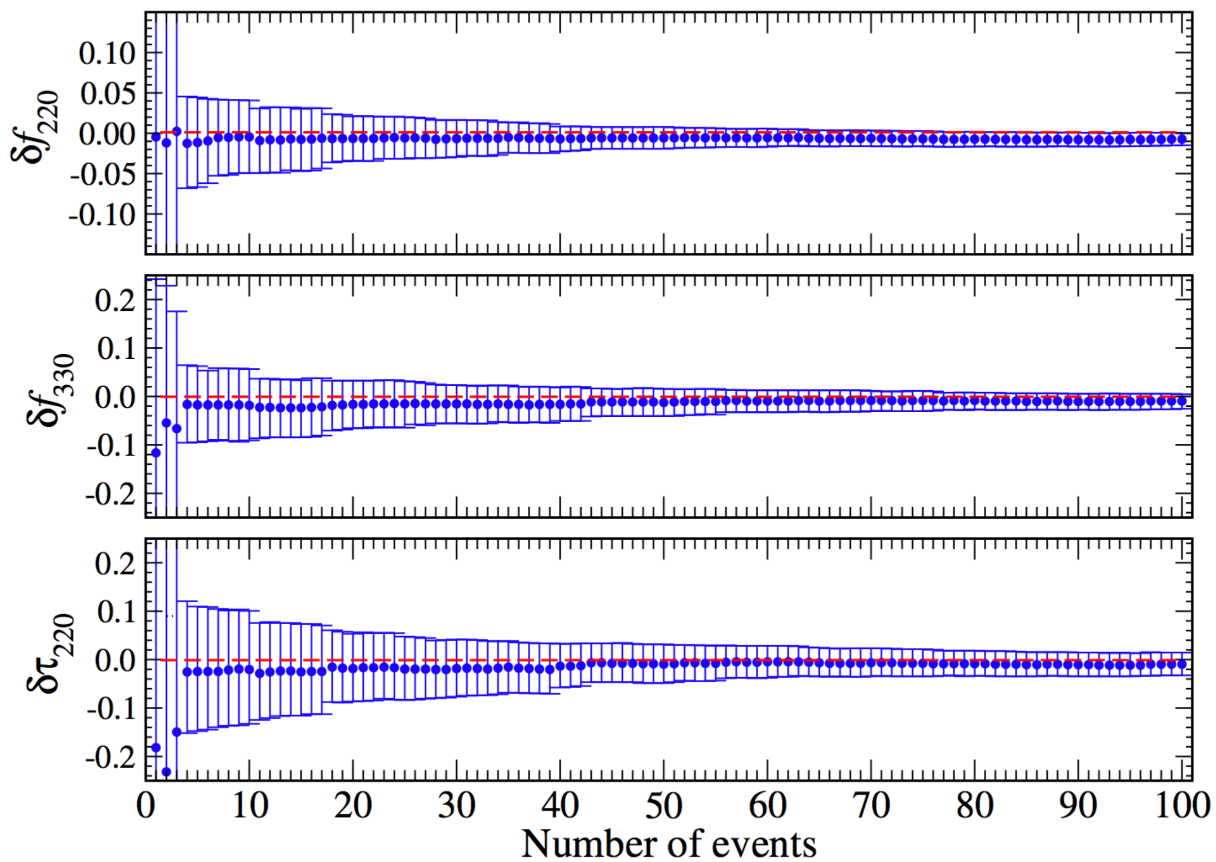


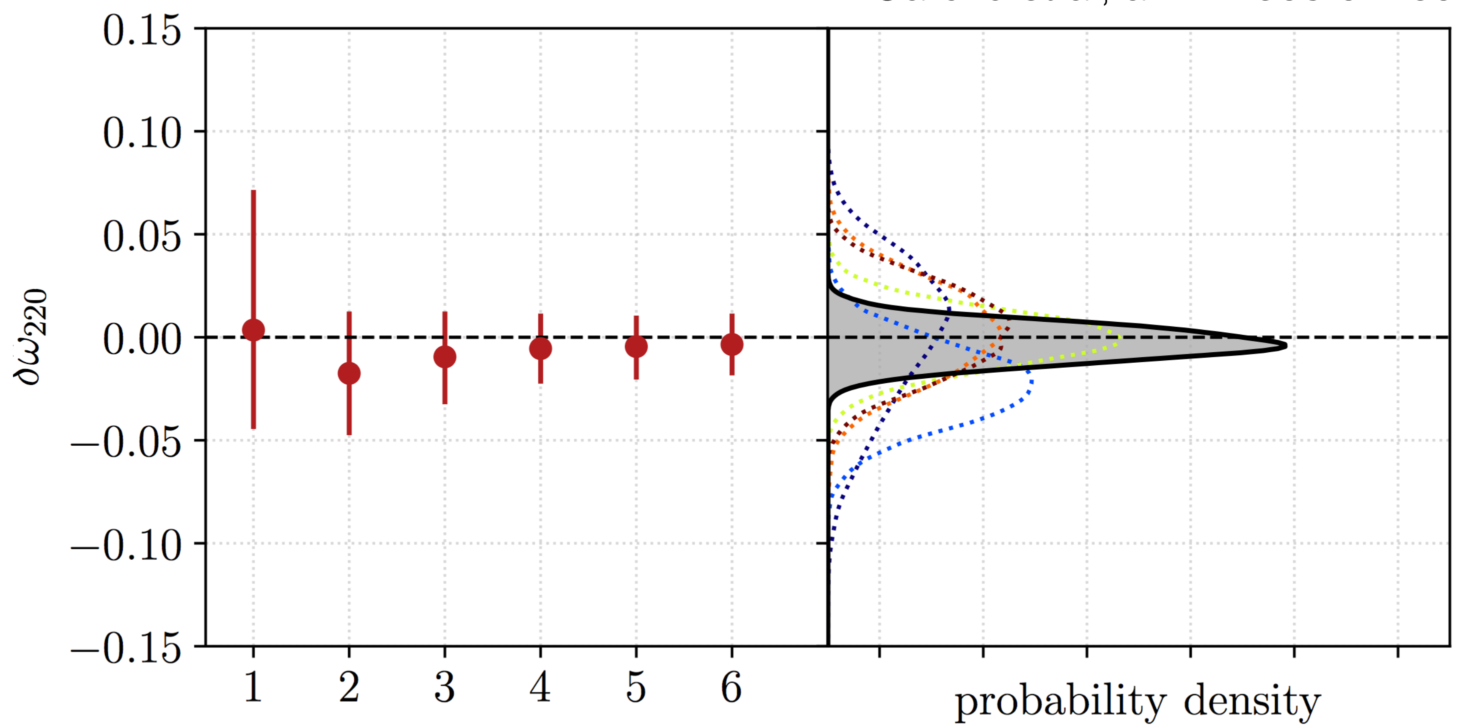
- Need at least two modes detected to test BH nature and "no-hair" theorem (e.g. Gossan et al, arXiv:1111.5819, Meidam et al,arXiv:1406.3201)

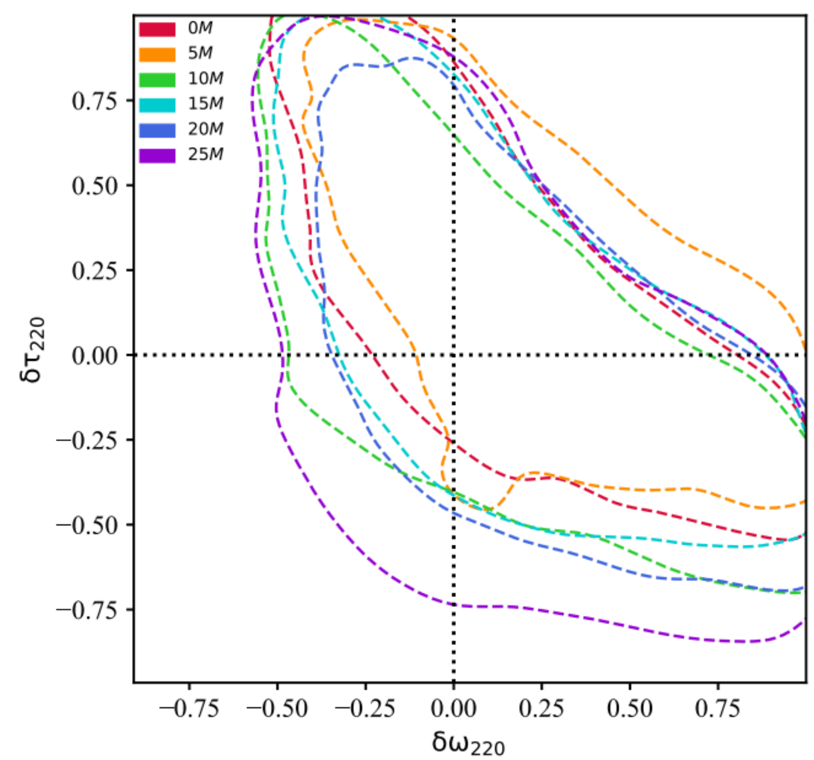
LVC,arXiv:1602.03841

Prospects for ringdown constraints

Carullo et al, arXiv:1805.04760



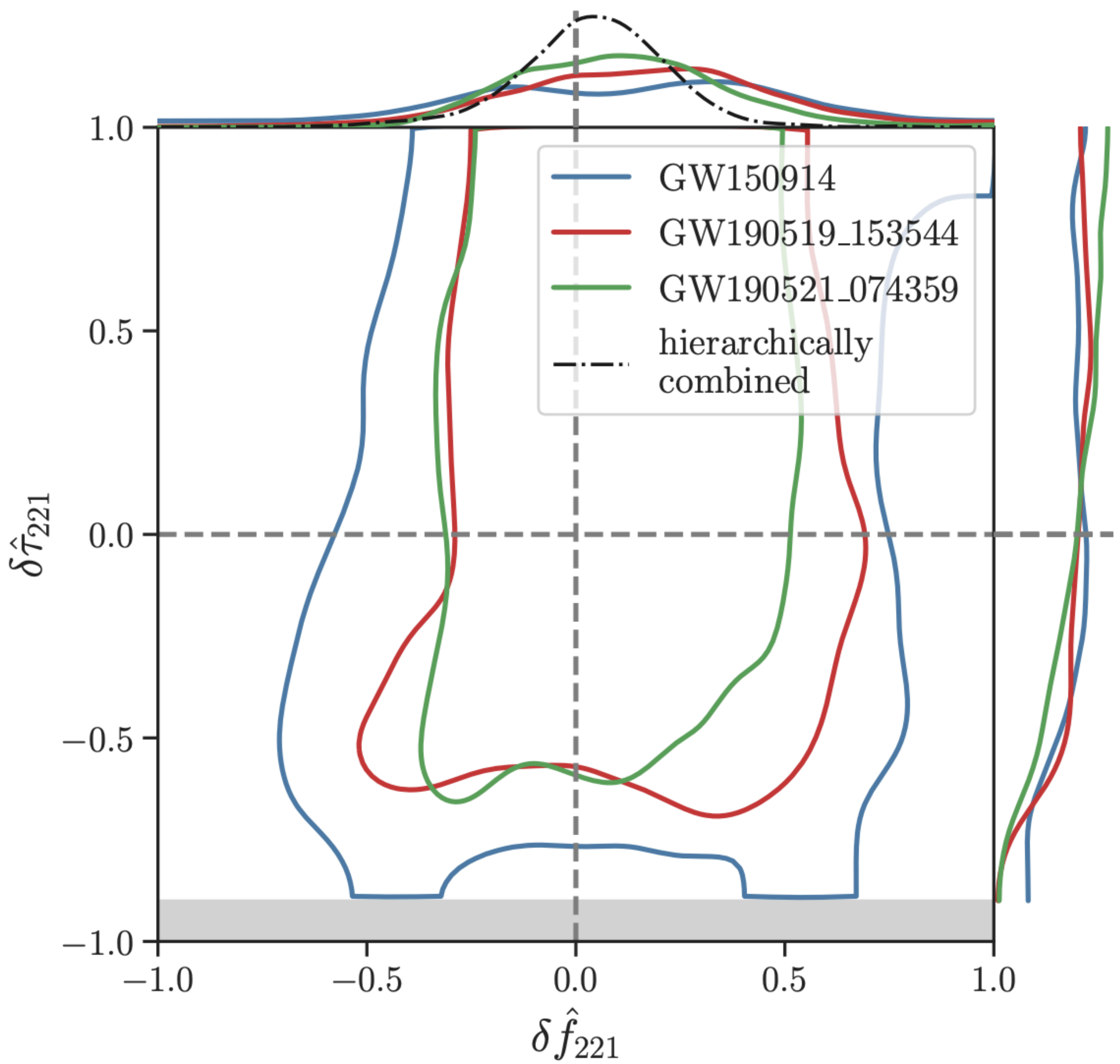


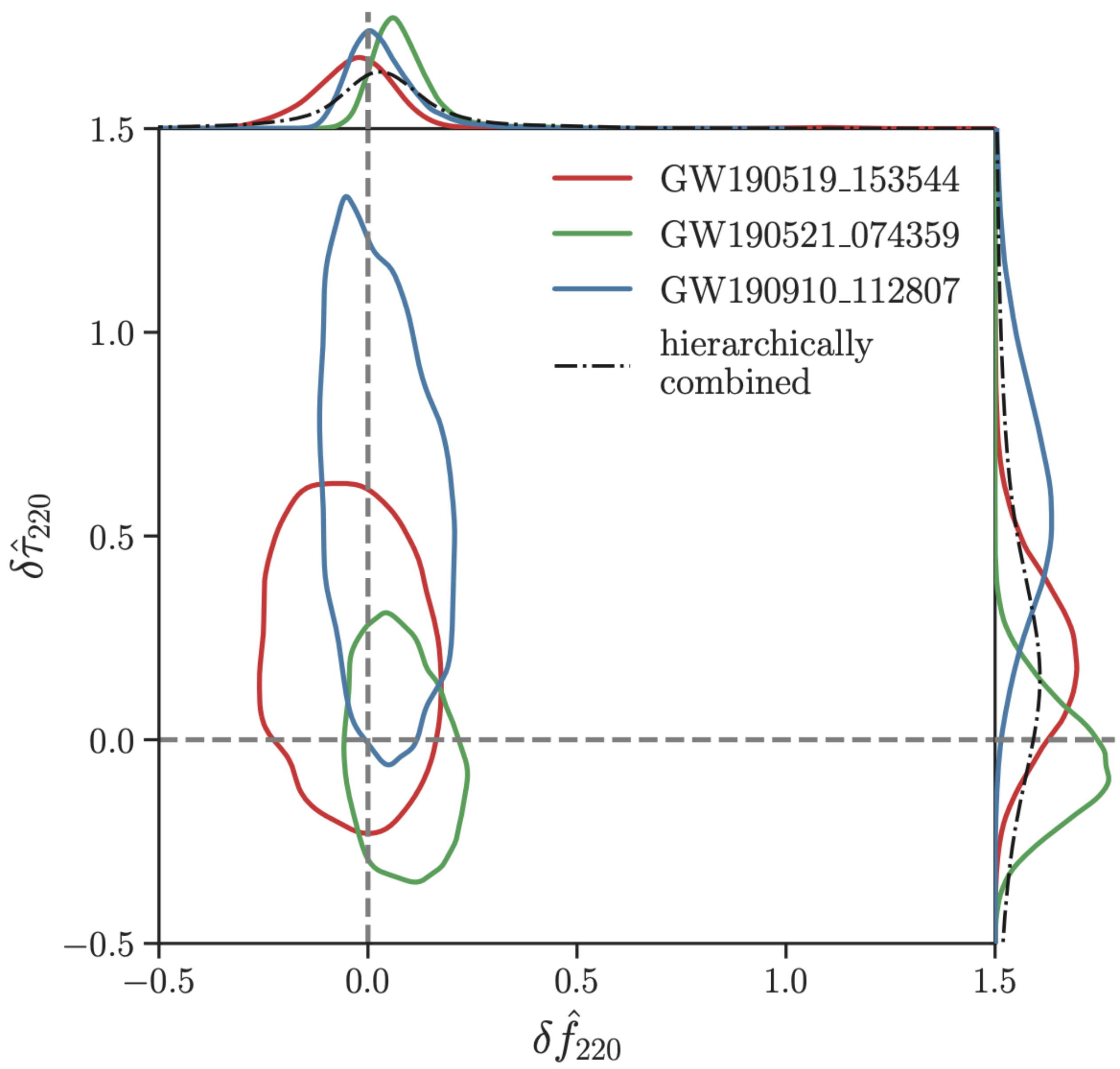


Brito et al, arXiv:1805.00293

Carullo et al, arXiv:1902.07527

Current ringdown constraints





LVC, https://dcc.ligo.org/public/0166/P2000091/010/o3a_tgr.pdf

Conclusions

- Gravitational wave astrophysics: the experimental study of the properties of dynamical space-time and black holes

  - Implications for nuclear physics (e.g. nuclear equation of state, interplay with sGRB, …)

  - Potential for studies of the large scale structure of the Universe

    - CBCs are tracers of the star formation and cosmic evolution

  - Access to the most extreme regions (known) in the Universe

Some useful references

- Sivia & Skilling, Data Analysis: A Bayesian Tutorial
- Jaynes, Probability Theory: The Logic of Science
- Gregory, Bayesian Logical Data Analysis for the Physical Sciences