Gravitational Waves I

Simone Mastrogiovanni
INFN Rome
ISCRA - Erice, Italy 2026

Istituto Nazionale di Fisica Nucleare
GravitySirens
Funded by the European Union
European Research Council
Established by the European Commission

What to expect?

Day 1: Introduction to Gravitational Waves (GW) and Data analysis techniques

- GW and interferometric detectors
- GW emitted by a compact binary coalescence (CBC) and the first evidence of GW emission
- Data analysis techniques for GWs: Frequentists VS Bayesians

Day 2: Physics and Astrophysics of notable Gravitational Wave sources

Day 3: Gravitational Wave Cosmology

- A review of the cosmic expansion
- Methodologies for GW cosmologya
- Current constraints

Gravitational Waves

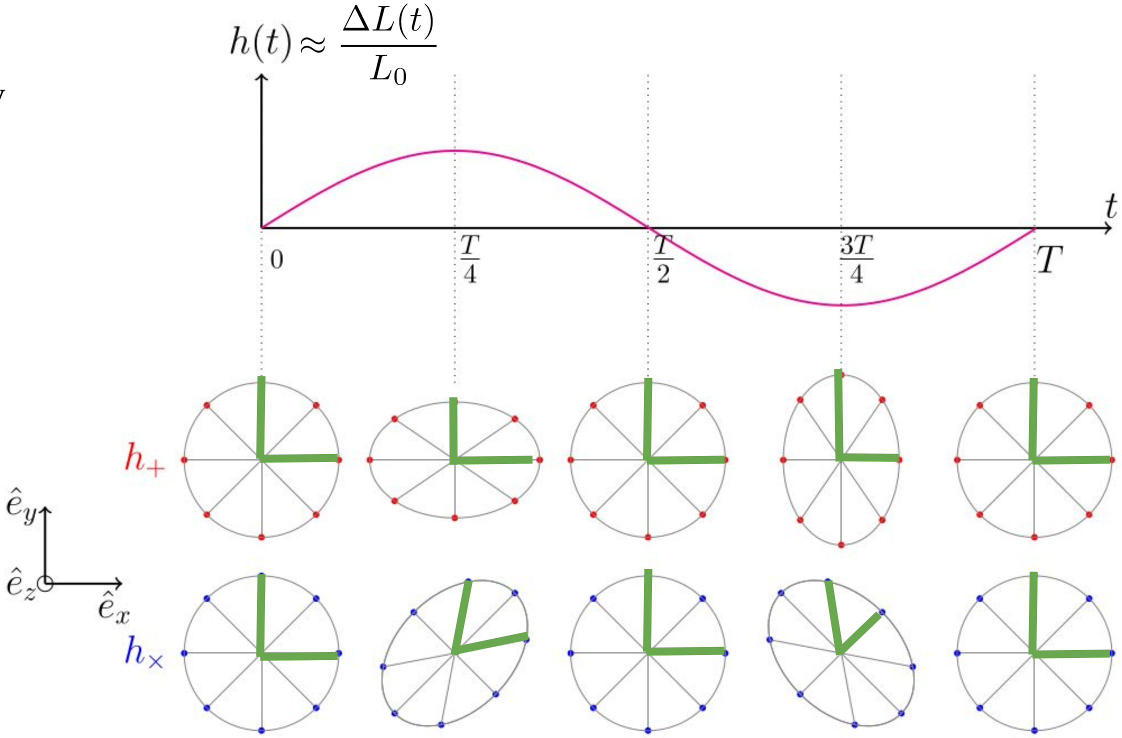
Gravitational Waves (GWs) are propagating perturbations in the spacetime metric caused by time-varying mass-energy distributions

GWs changes the proper distance between two freely-falling test masses and they have two polarizations.

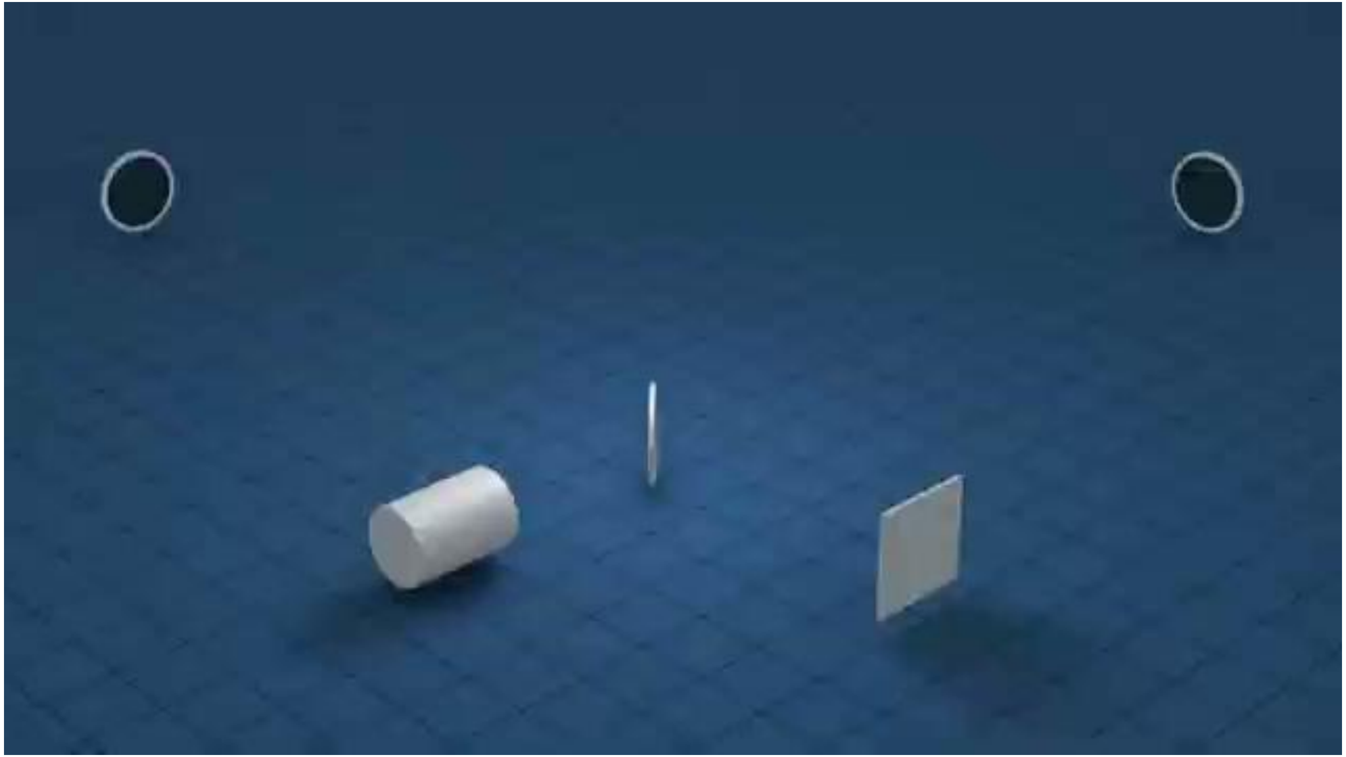
GW Quadrupole

$$h_{jk}^{\mathbf{TT}}(t,r) = \frac{2G}{c^4 r}\left[\frac{\mathrm{d}^2}{\mathrm{d}t^2} Q_{jn}^{\mathbf{TT}}(t-\frac{r}{c})\right]$$

$$\frac{8\pi G}{c^4} \simeq 10^{-43}\,\mathrm{kg}^{-1}\mathrm{m}^{-1}\mathrm{s}^{-2}$$

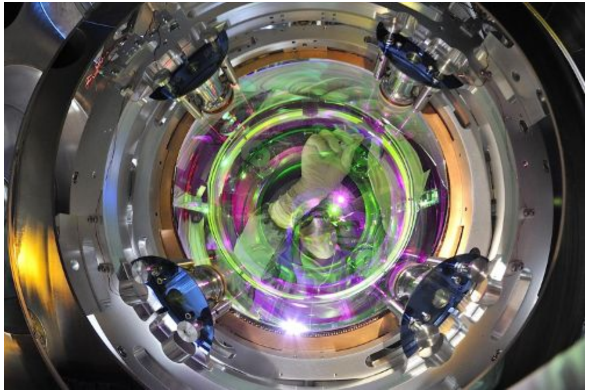


Gravitational Waves



Gravitational Waves

Superattenuators for seismic isolation



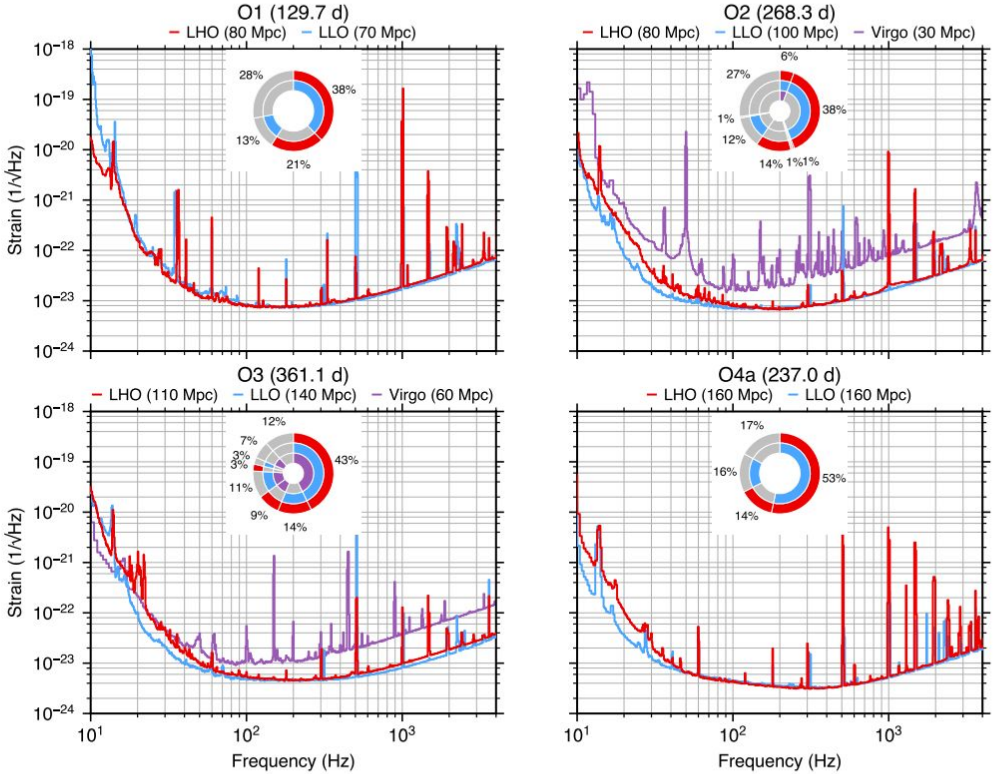
Mirrors weight 42 kgs to reduce the radiation presure noise

Coatings to reach a Finesse of ~400

Gravitational Waves

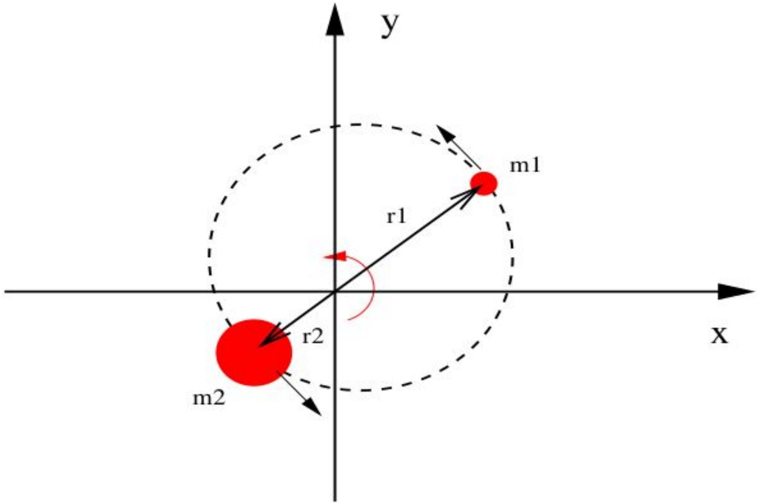
Gravitational Waves

- The sensitivity of a GW detector is estimated with the **Power Spectral Density** (PSD), namely the average fluctuation of the noise in a frequency range.
- LIGO, Virgo and KAGRA have progressively increased their sensitivity.
- Given the sources we are looking for, we can convert the PSD to a detection range.



LVK+ 2508.18080

GW emission from a binary



For a binary system in quasi-circular orbits GR predicts the emission of GWs.

The Luminosity (energy radiated) in GWs is

$$L_{\mathrm{GW}} \equiv \frac{\mathrm{d}E}{\mathrm{d}t} = \frac{1}{5}\frac{G}{c^5} \cdot \langle \dddot{I}_{jk} \dddot{I}_{jk} \rangle$$ **Quadrupole formula**

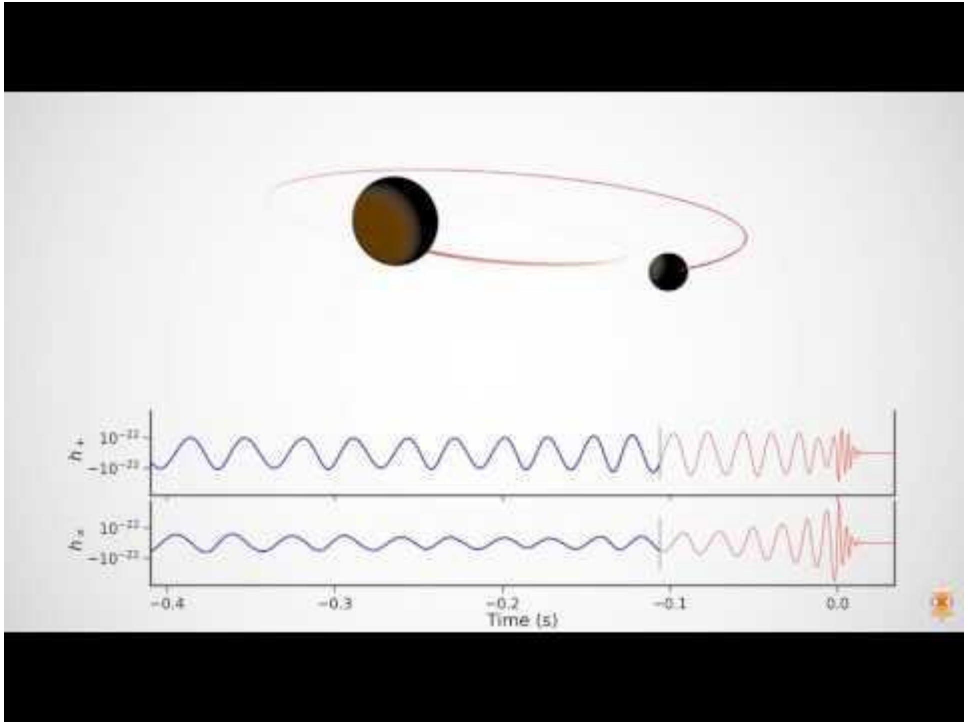
Where the I are the third derivatives w.r.t time of the quadrupole moment of the system.

For a system of point mass particles

$$I_{jk} = \sum_A m_A \left[x_j^A x_k^A - \frac{1}{3} \delta_{jk} \left(x^A\right)^2 \right]$$

GW emission from a binary

During the inspiral GWs are emitted and the system loses energy. The two objects rapidly approach each other.



GW emission from a binary

To understand how fast the two objects approach, we need to write the energy of the system. For a two-point mass particles.

Gravitational energy

$$E = -\frac{1}{2}\frac{G\mu M}{a}$$

Gravitational energy

$$M = M_1 + M_2$$

Reduced mass

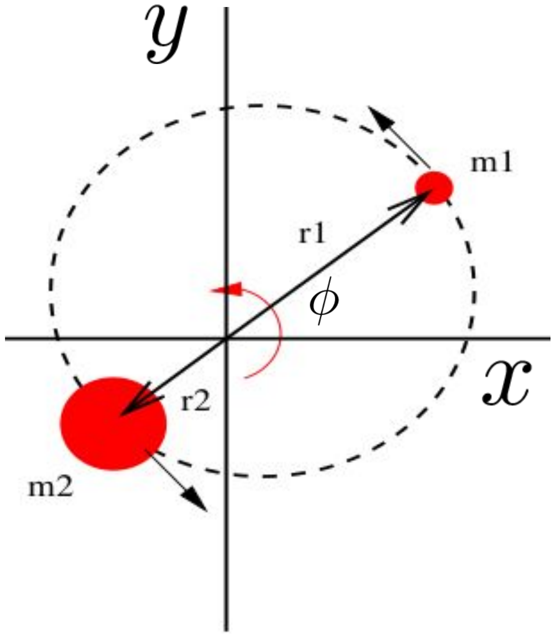
$$\mu = \frac{M_1 M_2}{M_1 + M_2}$$

We can derive the energy w.r.t to time to calculate the luminosity. By assuming that all the energy is emitted in GWs, we have

$$-\dot{E} = L_{\rm GW} = \frac{1}{2}\frac{G\mu M}{a^2}\dot{a} = -E \cdot \frac{\dot{a}}{a}$$

GW emission from a binary

$$I_{jk} = \sum_A m_A \left[x_j^A x_k^A - \frac{1}{3}\delta_{jk}\left(x^A\right)^2 \right]$$



XX- component

$$I_{xx} = \left(M_1 a_1^2 + M_2 a_2^2\right)\cos^2\phi + \text{constant terms}$$
$$= \frac{1}{2}\mu a^2 \cos 2\phi + \text{constant terms},$$

YY component

$$I_{yy} = -\frac{1}{2}\mu a^2 \cos 2\phi + \text{constant terms}$$

YX-XY component

$$I_{xy} = I_{yx} = \frac{1}{2}\mu a^2 \sin 2\phi + \text{constant terms}$$

GW emission from a binary

Adding all the components together, with more and more steps...

$$
\begin{aligned}
L_{\mathrm{GW}} &= \frac{1}{5}\frac{G}{c^5}\langle \dddot{I}_{jk}\dddot{I}_{kj}\rangle \\
&= \frac{1}{5}\frac{G}{c^5}\cdot(2\Omega)^6\cdot\left(\frac{1}{2}\mu a^2\right)^2\left(\sin^2 2\Omega t+\sin^2 2\Omega t+2\cos^2 2\Omega t\right) \\
&= \frac{32}{5}\frac{G}{c^5}\frac{(GM)^3}{a^9}\left(\mu a^2\right)^2 \\
&= \frac{32G^4}{5c^5}\frac{M^3\mu^2}{a^5}.
\end{aligned}
$$

GW emission from a binary

If the dynamic of the orbit is described by the Kepler law (radius of the orbit >> radius of the two bodies), then we can relate the shrinking of the orbit to the variation of the orbital period

Kepler's law

$$\Omega^2 = \frac{(2\pi)^2}{P^2} = \frac{GM}{a^3} \quad \Longrightarrow \quad \frac{\dot{P}}{P} = \frac{3}{2}\frac{\dot{a}}{a}$$

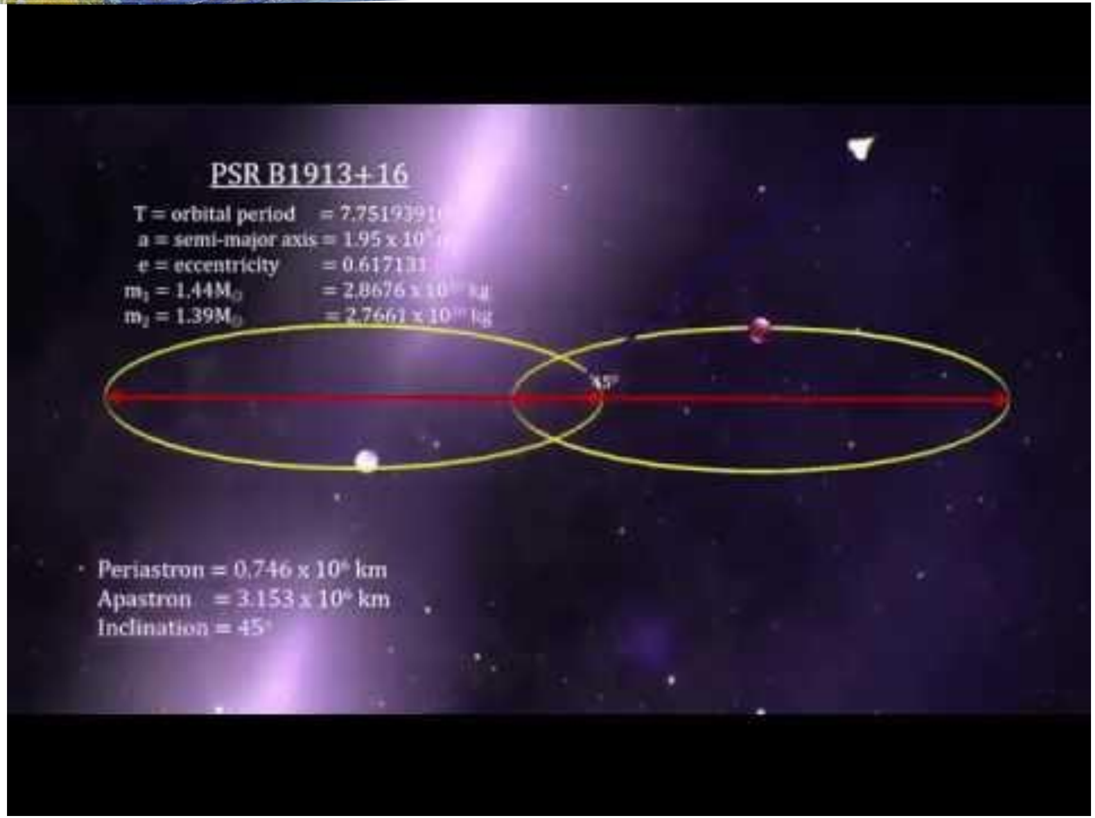
But the shrinking of the orbit can also be related to the GW luminosity

$$-\dot{E} = L_{GW} = \frac{1}{2}\frac{G\mu M}{a^2}\dot{a} = -E \cdot \frac{\dot{a}}{a} \quad \Longrightarrow \quad \frac{\dot{a}}{a} = +\frac{\dot{E}}{E}$$

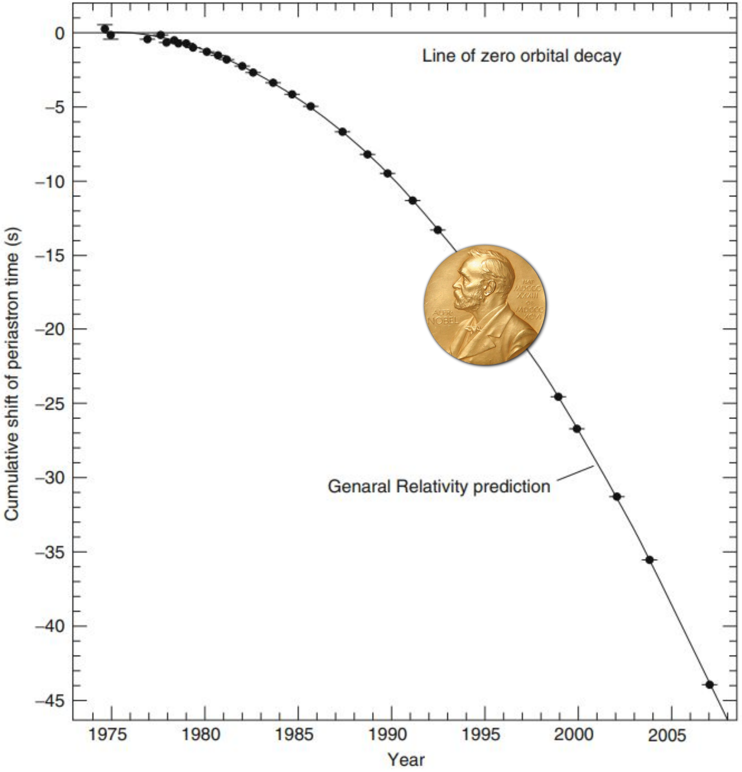
Therefore we can relate the orbital period variation to the GW luminosity

$$\frac{\dot{P}}{P} = \frac{3}{2}\frac{\dot{a}}{a} = +\frac{3}{2}\frac{\dot{E}}{E} = \frac{3}{2} \cdot \frac{32}{5}\frac{G^4}{c^5}\frac{M^3\mu^2}{a^5} \cdot \frac{(-2a)}{G\mu M} = -\frac{96}{5}\frac{G^3 M^2 \mu}{c^5 a^4}$$

GW emission from a binary



GW emission from a binary



Theory (taking into account ellipticity)

$$\frac{dP}{dt} = -2.4 \cdot 10^{-12}$$

Measured value

$$\frac{dP}{dt} = -(2.4184 \pm 0.0009) \cdot 10^{-12}$$

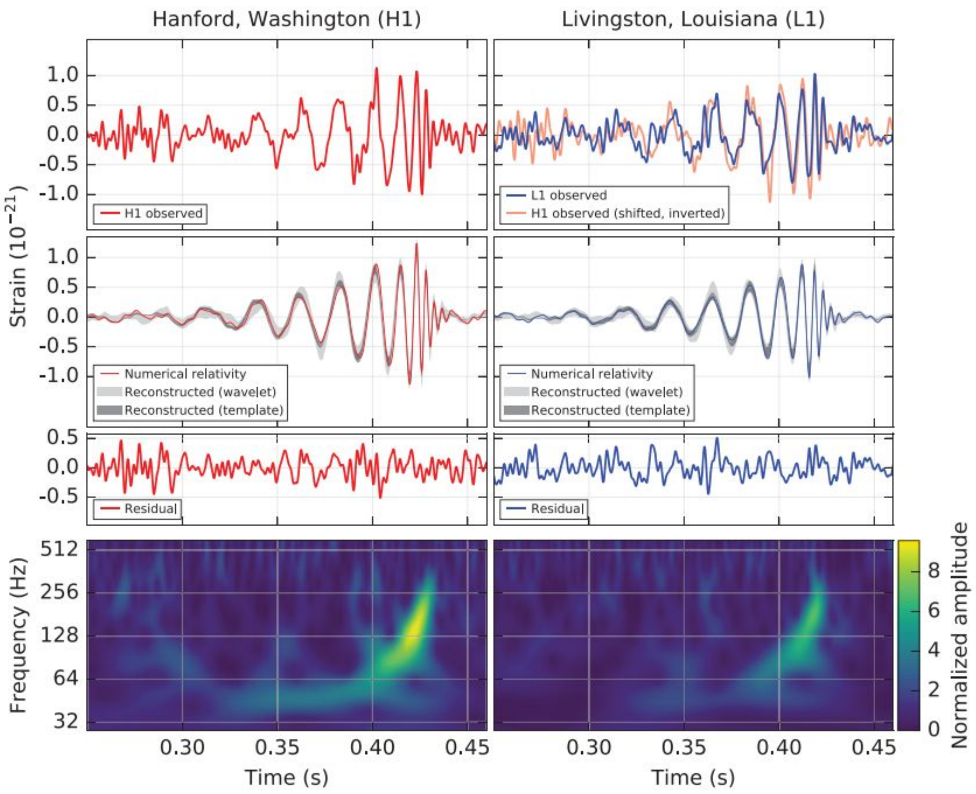
Good agreement between theory and observations.

In the following years this kind of measure has been repeated for other binary pulsars (PSR J0737-3039A)

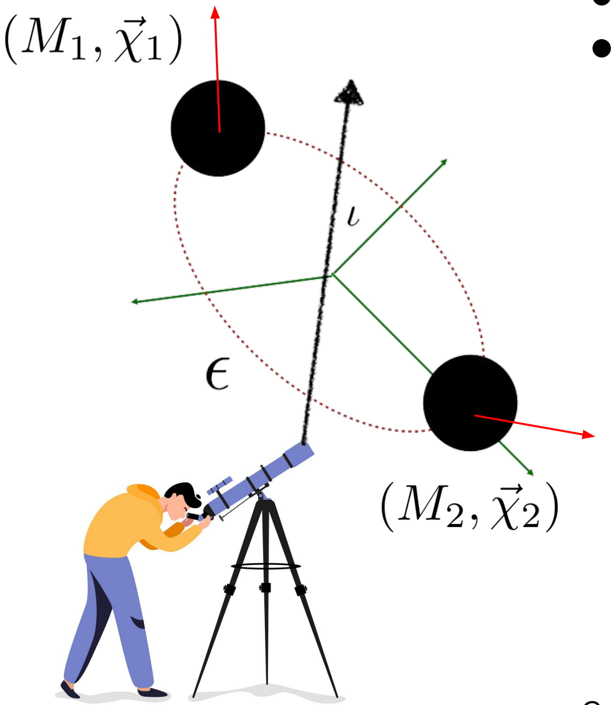
Gravitational Waves

The first direct detection of GWs

- GW150914 was detected on September 14, 2015 at 09:50:45 UTC.
- The signal was observed between 35 and 250. The strain peak was around 1e-21.
- The signal had a SNR of 24.

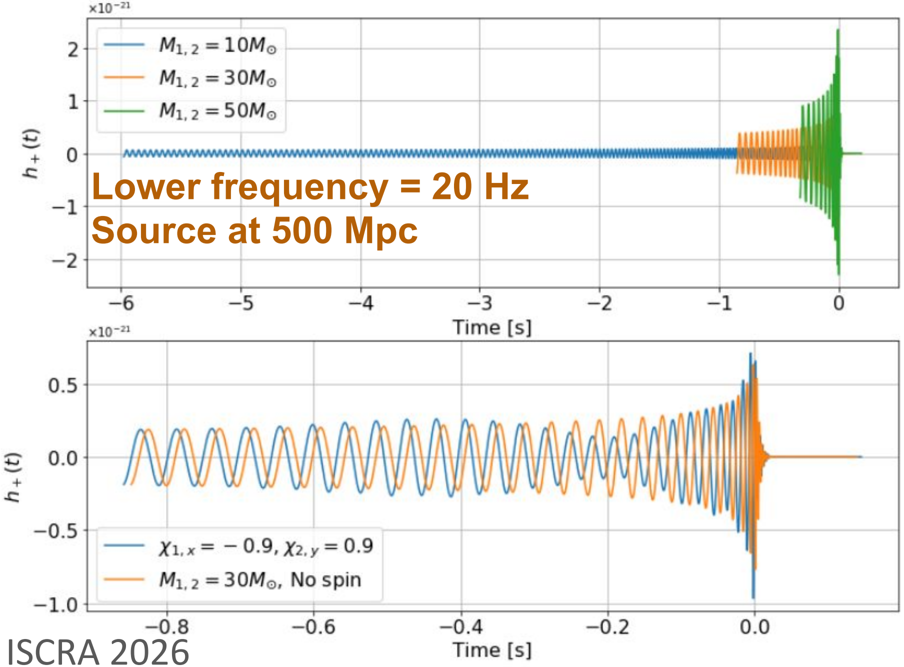


The first direct detection of GWs



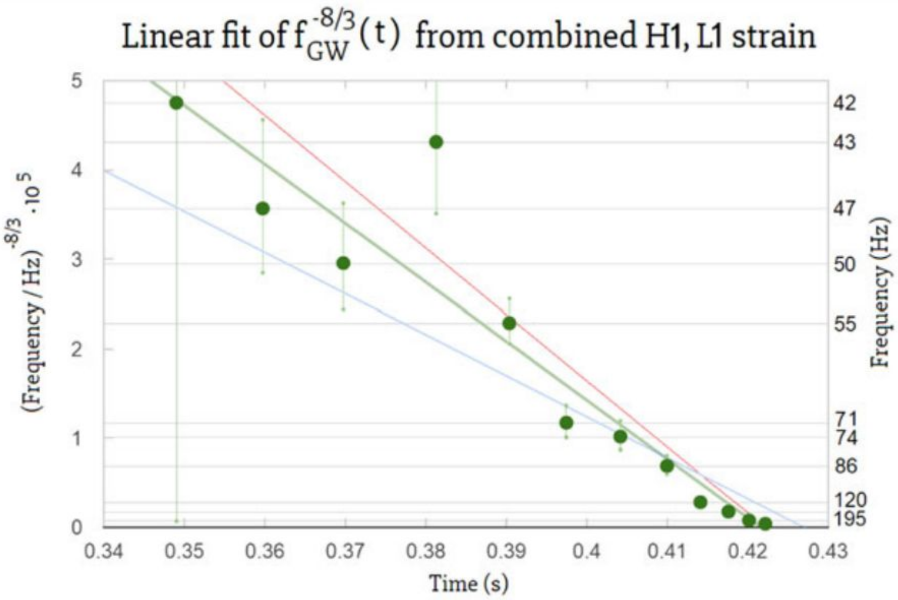
The parameters:

- **Intrinsic:** Spins, Masses, tidal deformability, Eccentricity
- **Extrinsic:** Time, reference phase, sky position, luminosity distance, orbital orientation



Generated with pycbc

The first direct detection of GWs



We can estimate the source-frame mass of the system by remembering that the chirp mass is related to the frequency evolution.

$$f_{GW}^{-8/3}(t) = \frac{(8\pi)^{8/3}}{5}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/3}(t_c - t)$$

The chirp mass can be estimated from the previous plot doing a linear fit.

If you want to try it:

- Check for the paper describing this activity.
- The frequency of the GW is saved in this file.

The first direct detection of GWs

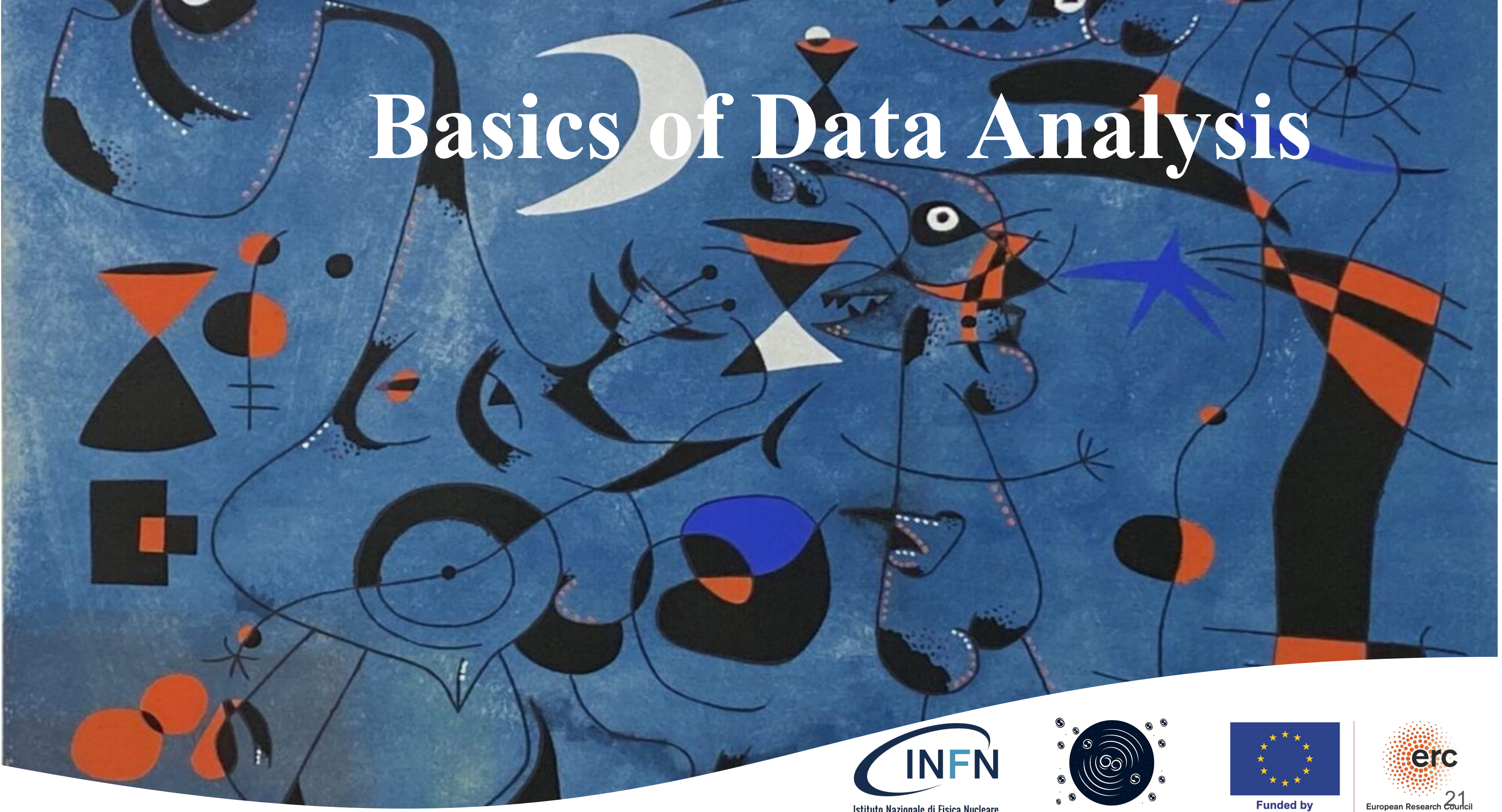
Once you have fit for the chirp mass, by assuming that the two masses are equal you will see that

$$m_1 = m_2 = 2^{1/5}\mathcal{M} = 35\,\mathrm{M}_\odot \qquad M = m_1 + m_2 = 70\,\mathrm{M}_\odot$$

We can then convert the total mass to the radius of the orbit using the Kepler's law. You will see that the radius of the orbit is of the same order of the Schwartchild radius of the two objects.

$$R = \left(\frac{GM}{\omega^2_{\mathrm{Kep}}\big|_{\max}}\right)^{1/3} = 350\,\mathrm{km} \qquad r_{\mathrm{Schwarz}}(m) = \frac{2Gm}{c^2} = 2.95\left(\frac{m}{\mathrm{M}_\odot}\right)\mathrm{km}$$

Basics of Data Analysis



Istituto Nazionale di Fisica Nucleare
GravitySirens
Funded by the European Union
European Research Council
Established by the European Commission

A few references

- Gravitational Wave Open Data workshop: Online course with lectures, Tutorials.

- A good introduction to GW DA aspects (from the GW stochastic searches)

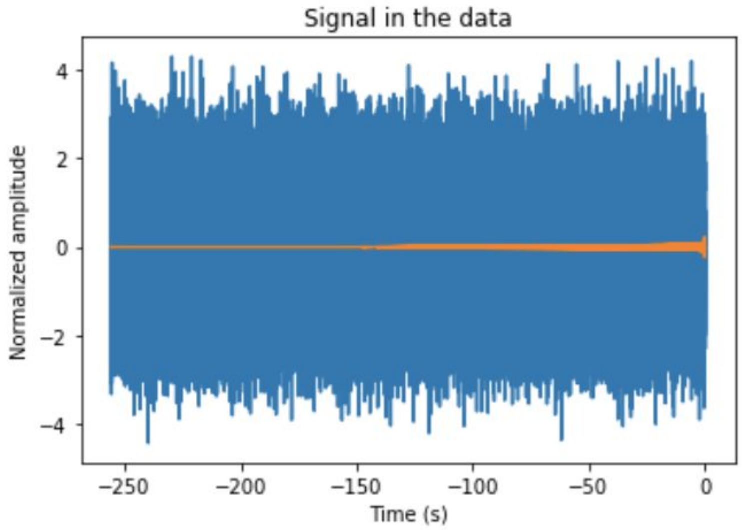
- An introduction to LVK CBC signal extraction (from C. Berry blog article).

- Information Theory, Inference, and Learning Algorithms - D. MacKay (for a longer introduction)

A brief introduction to DA

The scope of data analysis for GWs:

- **Is there a signal?**
  - Detection problem
- **What parameters the signal has?**
  - Parameter estimation
- **Is it a Binary Neutron star or black hole?**
  - Model selection problem



In this data strain there is a Binary Neutron star signal with 1.4-1.4 solar masses.

Introduction to LVK DA

The data is the composed by:

- Noise: We assume the noise to be gaussian and coloured noise (we will see later what this means.)

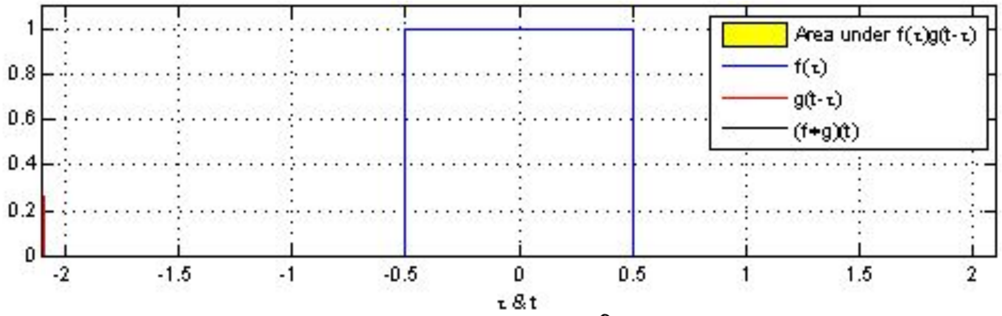
- The signal from the GW source reprocessed by the detector response.

$$d = \{n_1 + F[h_1(\theta)], \ldots, n_N + F[h_N(\theta)]\}$$

The data can be though as a collection of data points with signal and noise.

The convolution

- We want to look for this signal with “matched filtering”.
- Matched filtering cross-correlate your data with a template.
- When the template matches the signal in data, then you have a spike in the cross-correlation



$$C(\tau) = g(t) \star f(t) = \int_t f(t-\tau)g(t)dt$$

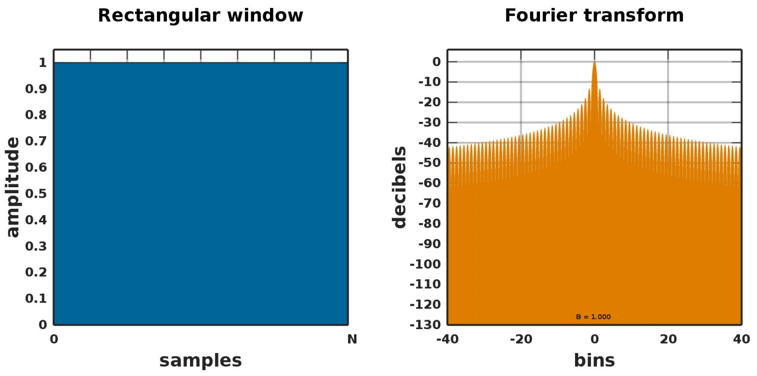
The convolution theorem

The convolution theorem relates multiplication of functions in time/frequency domain to convolutions in frequency/time domain

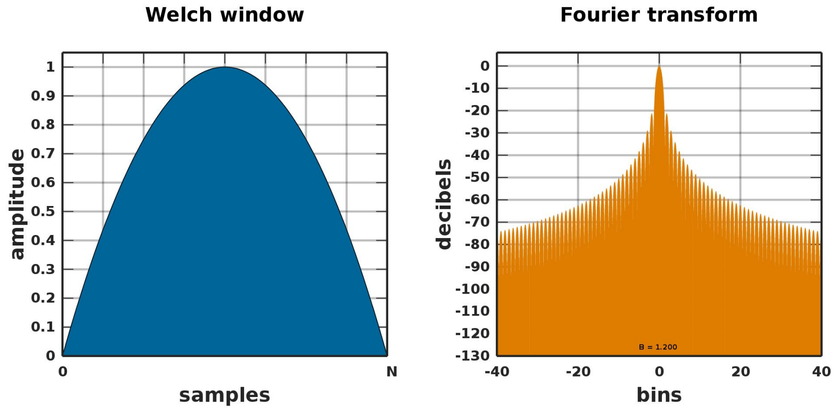
$$\mathcal{F}[\int_t f(t-\tau)g(t)dt] = \tilde{f}(\omega)\tilde{g}(\omega)$$

$$\mathcal{F}^{-1}[\int_\omega \tilde{f}(\omega-\omega')\tilde{g}(\omega)d\omega] = f(t)g(t)$$

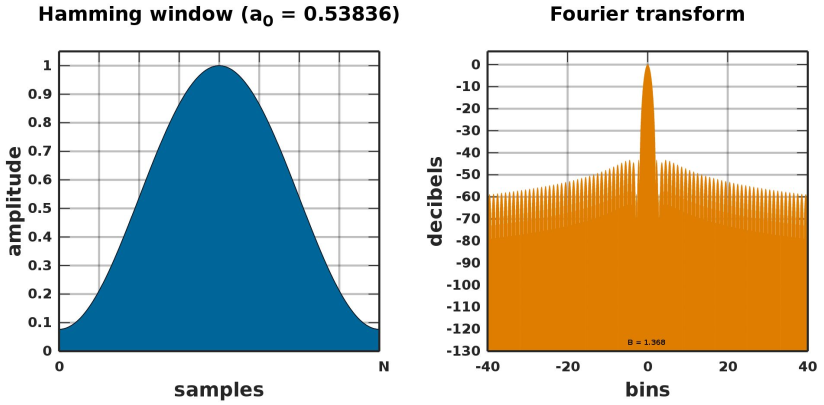
Windows



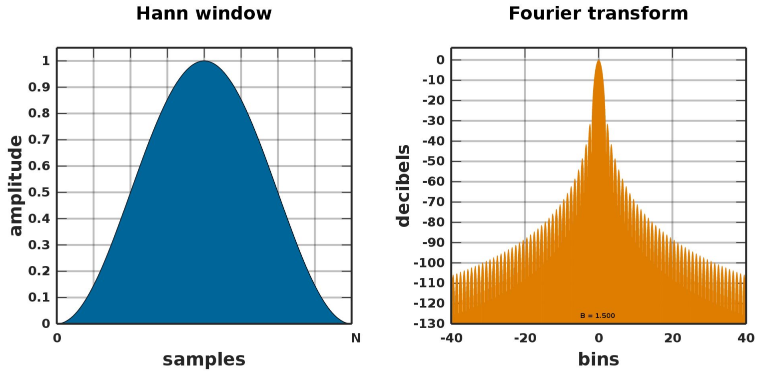
Rectangular window
amplitude
samples
0
N
Fourier transform
decibels
bins
B = 1.000



Welch window
amplitude
samples
0
N
Fourier transform
decibels
bins
B = 1.200



Hamming window (a_0 = 0.53836)
amplitude
samples
0
N
Fourier transform
decibels
bins
B = 1.368

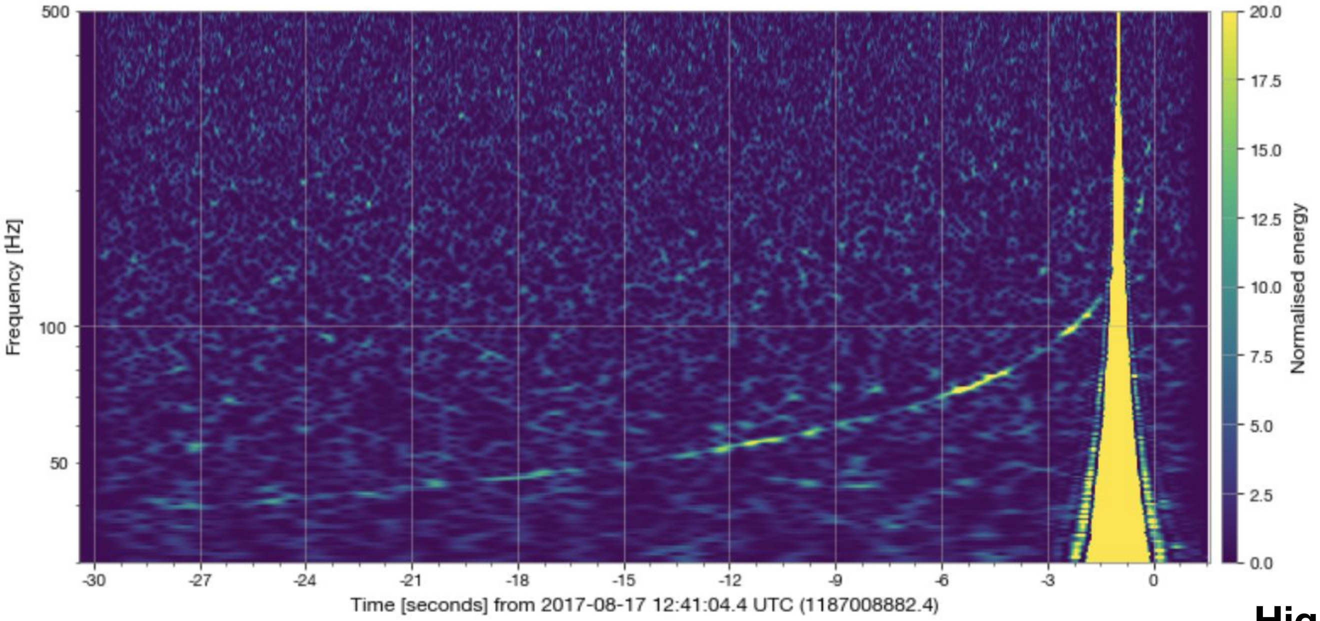


Hann window
amplitude
samples
0
N
Fourier transform
decibels
bins
B = 1.500

The Q-transform

Here we are dividing the data in 4 seconds chunks and for each of them we plot and FFT.

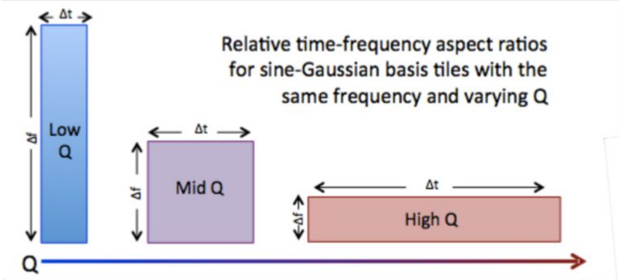
Higher time resolution

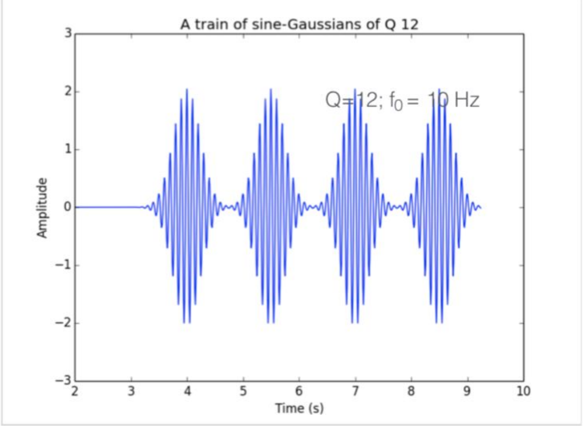


Higher frequency resolution

The Q-transform

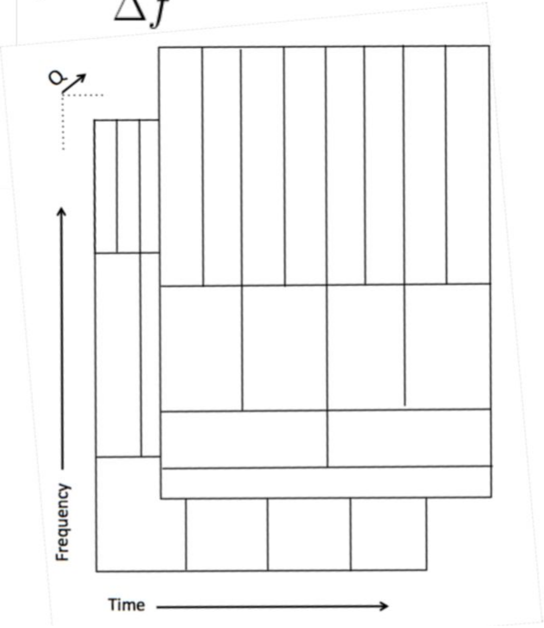
Time-Frequency plots for LIGO are generated with a special FT.





S. Chatterji et al. CQG (2010)
Images: McIver

$$Q = \frac{f_0}{\Delta f}$$



Introduction to matched filtering

$$d(t) = n(t) + h(t)$$

We want to find the optimal filter to convolve with the data $f(t)$

$$c(t) = d(t) \star f(t) = n(t) \star f(t) + h(t) \star f(t)$$

We want to find the optimal filter that maximise the following quantity(SNR)

$$\mathrm{SNR}^2 = \frac{|h(t) \star f(t)|^2}{E[|n(t) \star f(t)|^2]}$$

We whitened the noise

Introduction to matched filtering

$$d(t) = n(t) + h(t)$$

We want to find the optimal filter to convolve with the data $f(t)$

$$c(t) = d(t) \star f(t) = n(t) \star f(t) + h(t) \star f(t)$$

We want to find the optimal filter that maximise the following quantity (SNR)

$$\mathrm{SNR}^2 = \frac{|h(t) \star f(t)|^2}{E[|n(t) \star f(t)|^2]} = \frac{|\int h(f) F(f) e^{2\pi i f \tau} df|^2}{E[|\int n(f) F(f) e^{2\pi i f \tau} df|^2]}$$

Introduction to matched filtering

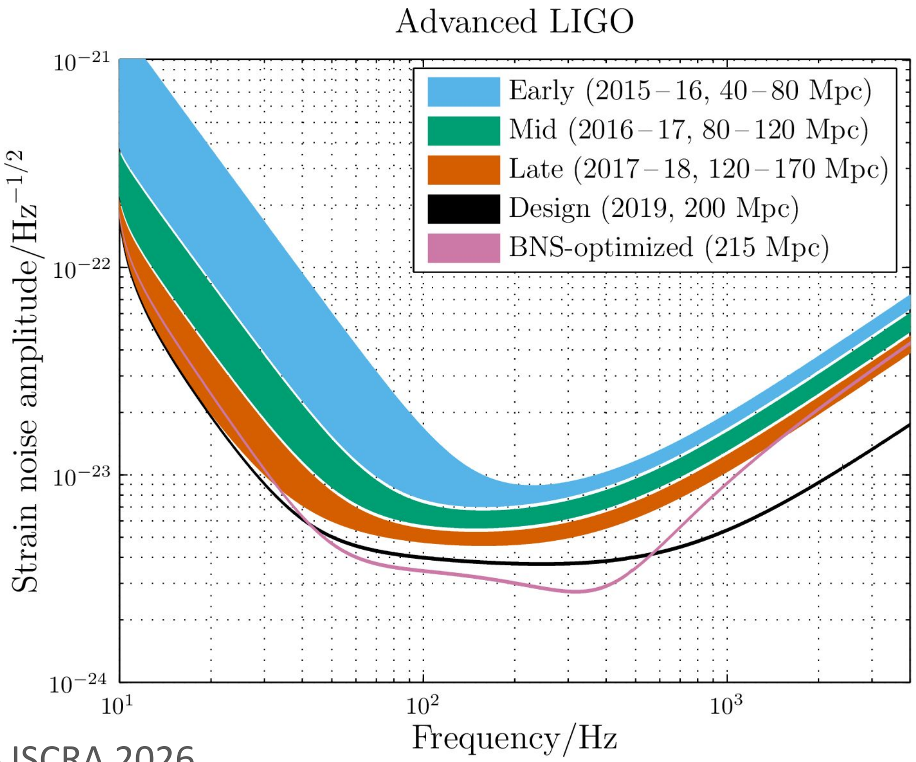
The noise background that you have seen in the previous plots it is called power spectral density (PSD)

Power spectral density connected to the fluctuation of gaussian noise.

$$p(n) \propto e^{-\frac{1}{2} n_i C_{ij}^{-1} n_j}$$

$$C_{i,j} = \frac{1}{2} S_n(f_i) \delta_{ij}$$

Can have different amplitude at different frequencies.



Introduction to matched filtering

Remember from the previous slide that: $E[|n(f)|]^2 = S_n(f)$

$$\frac{|\int h(f)F(f)e^{2\pi i f\tau}df|^2}{\int S(f)|F(f)|^2 df} \leq \frac{|\int |h(f)|^2 df \int |F(f)|^2 df}{S(f)\int |F(f)|^2 df}$$

I have assumed that the spectrum is “flat” and used the Cauchy-Schwartz inequality

The Matched filter is the one that allow me to write the Equality in the above disequality, namely

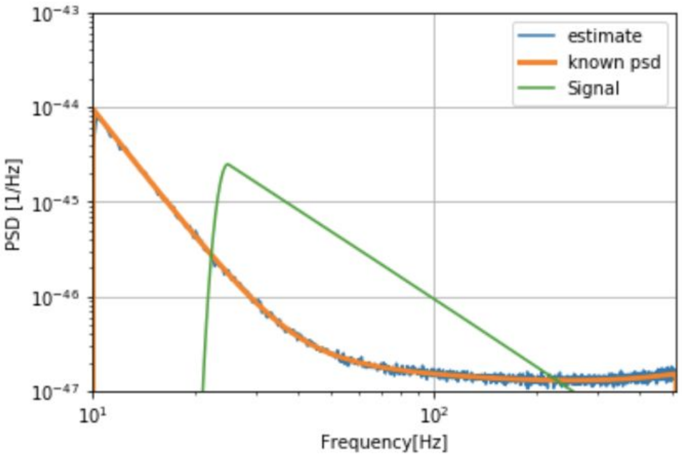
$$F(f) = h(f)e^{-2\pi i f\tau}$$

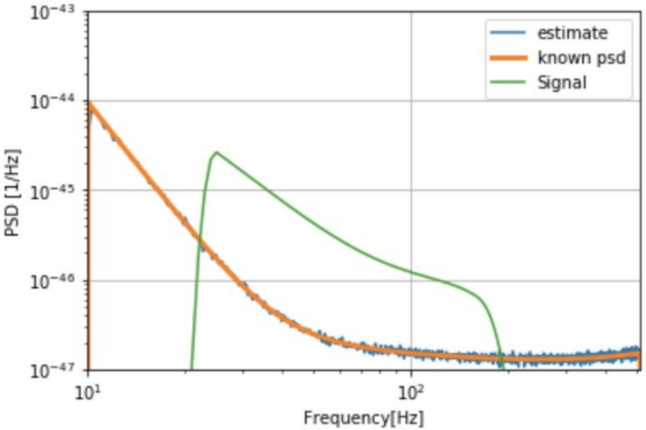
Introduction to matched filtering

With these choice of filter, the SNR become

$$SNR^2 \equiv < h|h >= \int_{f_l}^{f_u} \frac{|h(f)|^2}{S_n(f)} df$$

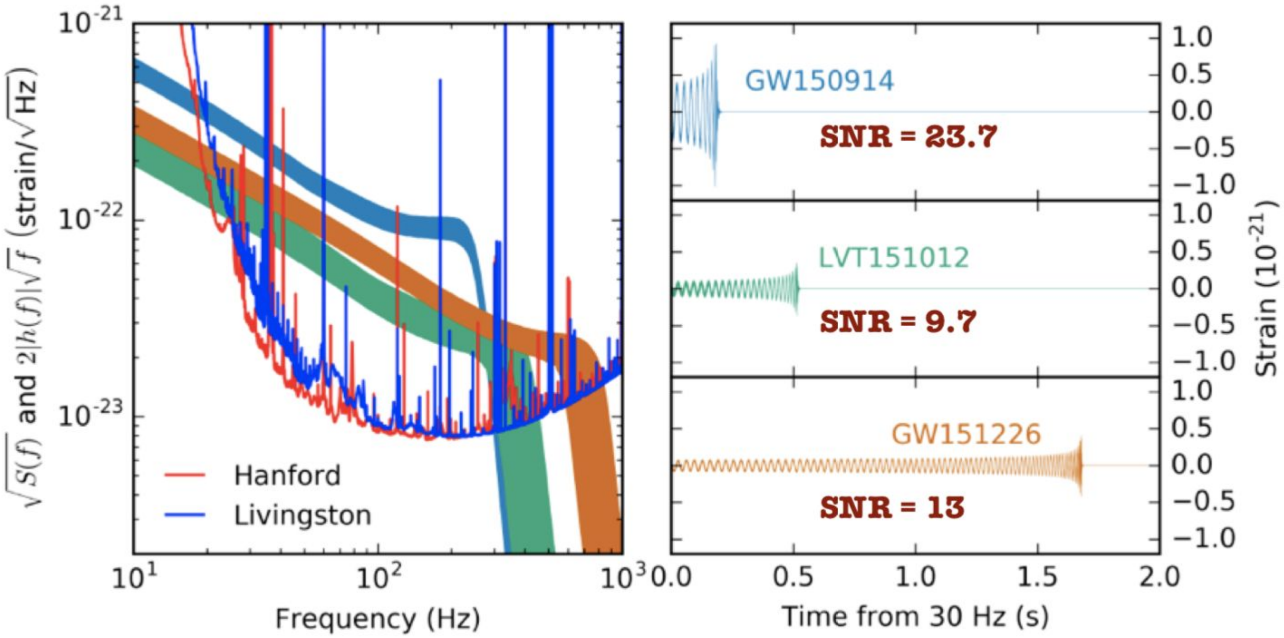
The SNR is basically an area!





Introduction to matched filtering

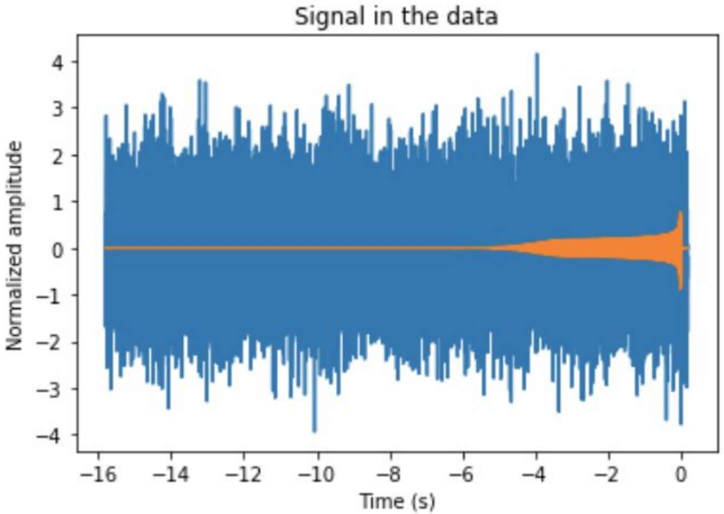
Signal to noise events of several events at comparison

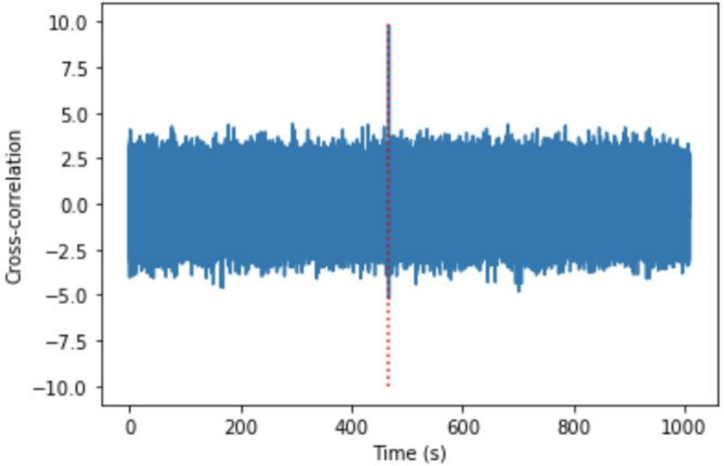


Introduction to matched filtering

1. We take calculate a template for the waveform, corresponding to some physical parameters, e.g. masses etc.

2. We slide it on the data and look for excess in the cross-correlation/SNR.



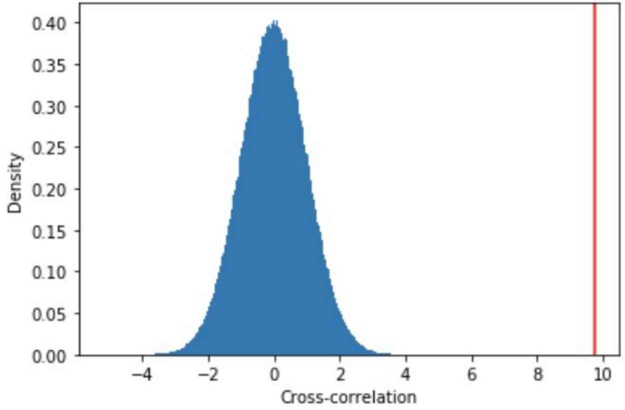


Introduction to matched filtering

1. When the template is matching the noise, we generate *noise background* realizations, that can be used to assess the significance of our candidate.

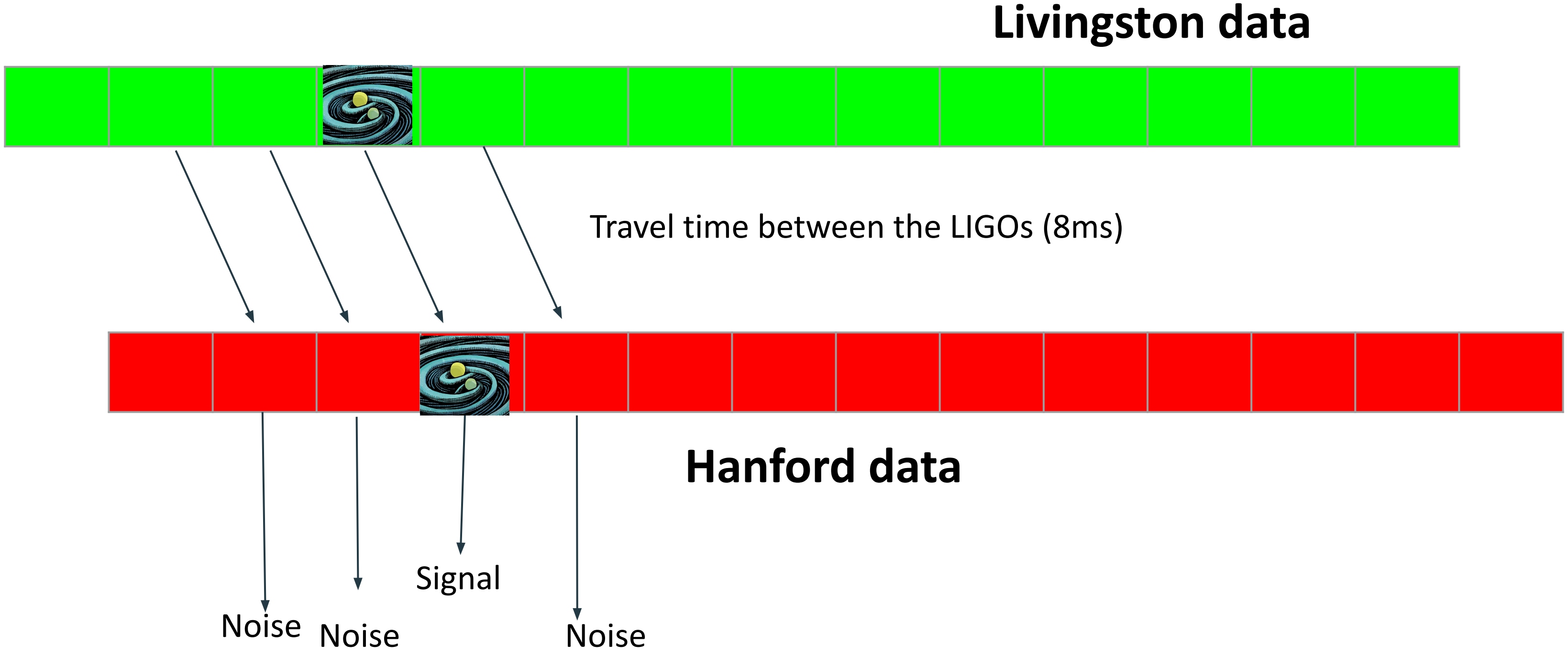
2. When we match the signal, we obtain a very strong outlier, which is not compatible with the background distribution of the noise.

3. A preliminary significance can be calculated using the p-value (or False alarm probability).



Introduction to matched filtering

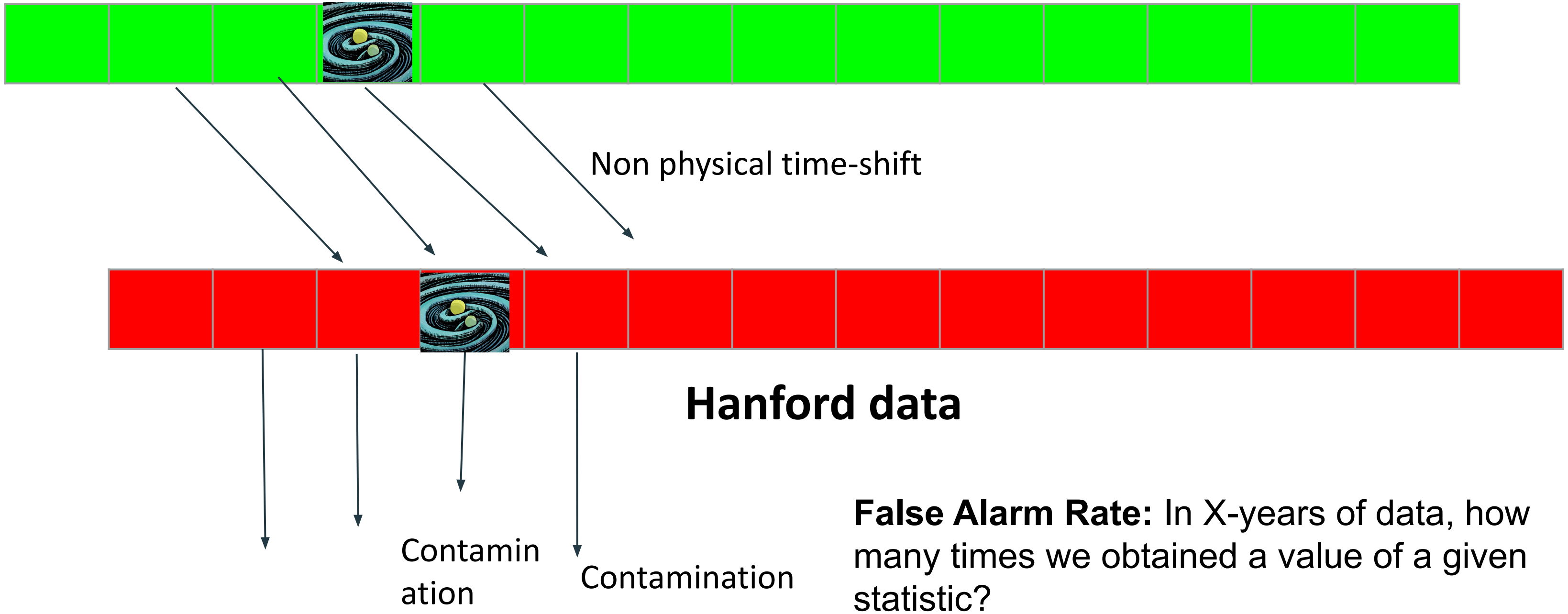
In a real search, we take all of our data and we slide pre-built templates for each interferometer. Then we match the interferometer results taking into account the travel-time between them.



Introduction to matched filtering

To generate noise backgrounds we perform the same search but matching the data between interferometers ``*wrongly*``

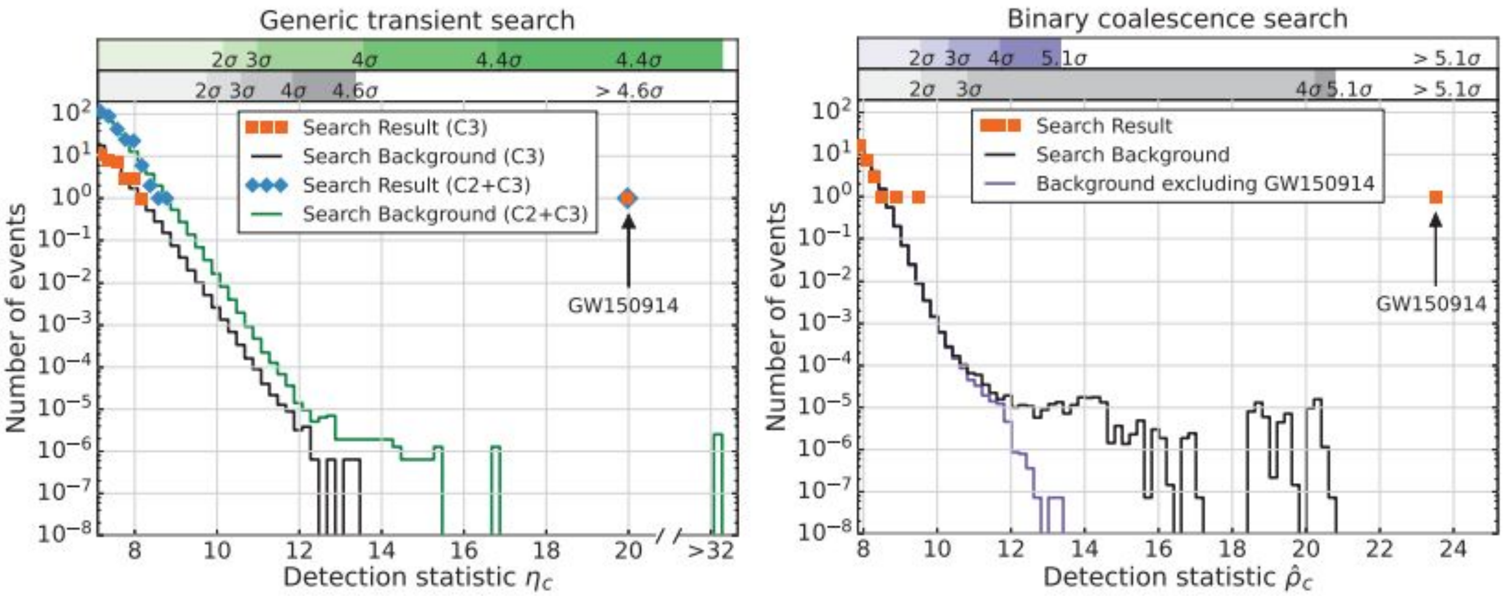
Livingston data



False Alarm Rate: In X-years of data, how many times we obtained a value of a given statistic?

Introduction to matched filtering

- The significance of this event was 1 event per 203 000 years, equivalent to a significance greater than 5.1σ.
- The event was so loud that contaminated the noise backgrounds when time-sliding data.



The Q-transform

Hands-on session: Spectrograms, spectra and Q-transform

Link

Introduction to matched filtering

If our noise is gaussian with 0 mean and given PSD, then the likelihood for obtaining N points of noise is

$$p(n) \propto e^{-\frac{1}{2} n_i C_{ij}^{-1} n_j}$$

$$C_{i,j} = \frac{1}{2} S_n(f_i)\delta_{ij}$$

However, remember that your data is a super-position of noise and signal… hence

$$p(n) = p(d-h) \propto e^{<d-h|d-h>}$$

Where I have defined

$$< a|b >= 2\int_{f_l}^{f_u} \frac{a(f)b^*(f) + a^*(f)b(f)}{S_n(f)} df$$

Introduction to matched filtering

You can verify that subtracting the signal in data gives you back

$$p(n) \propto e^{-\frac{1}{2} n_i C^{-1}_{ij} n_j}$$

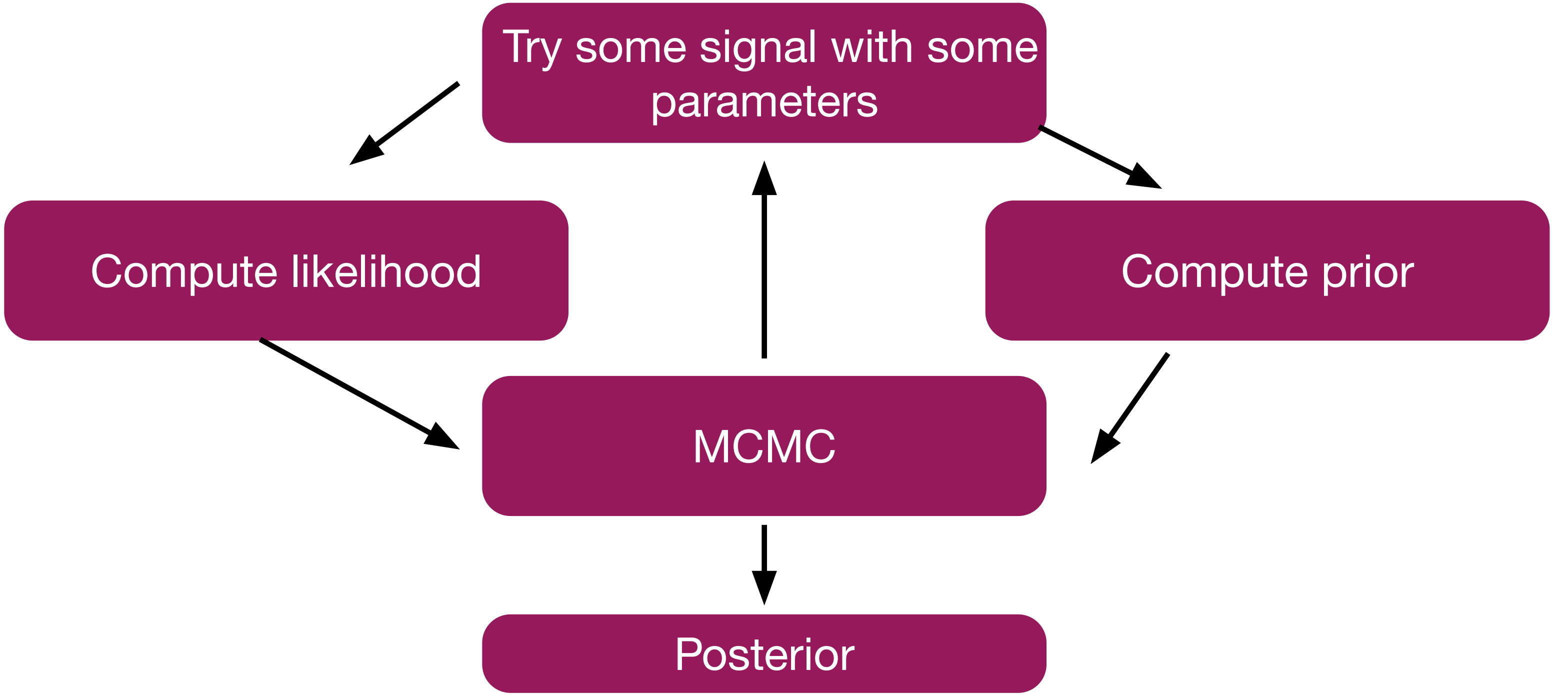
The scalar product that we have introduced, is directly related to the definition of the SNR. In fact, as we have seen before.

The higher the SNR, the more peaked is the likelihood

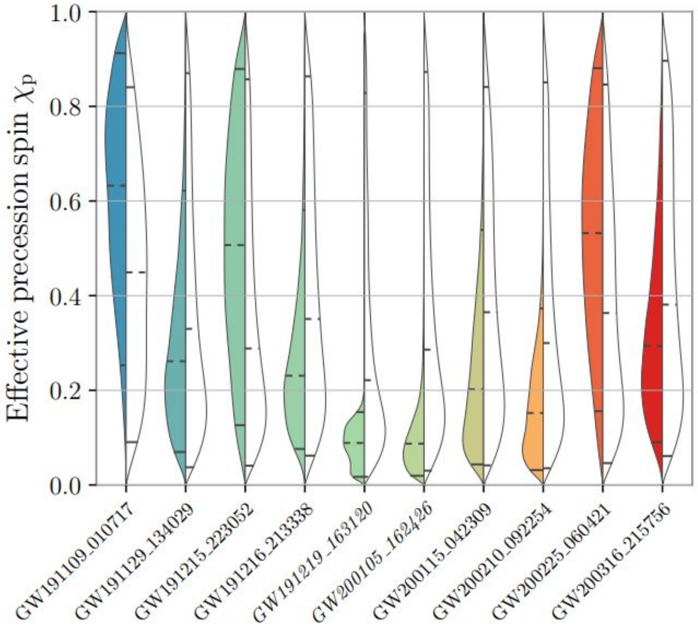
$$SNR^2 \equiv < h|h >= \int_{f_l}^{f_u} \frac{|h(f)|^2}{S_n(f)} df$$

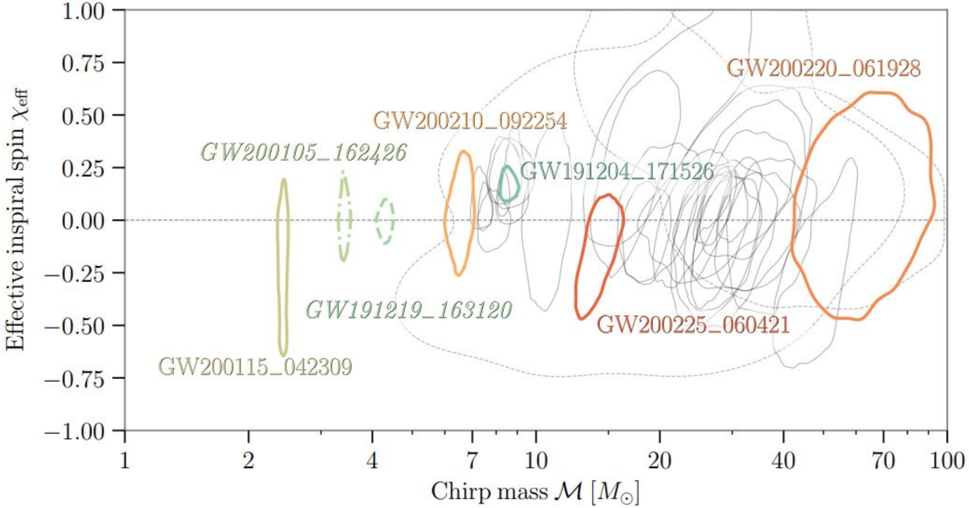
Introduction to matched filtering

Procedure that we follow in GW data analysis for estimating the parameters of a signal



Introduction to matched filtering





Part 1 - Extra slides

INFN
Istituto Nazionale di Fisica Nucleare
GravitySirens
Funded by
the European Union
erc
European Research Council
Established by the European Commission

A brief introduction to DA

We have two possible kind of approach to data analysis

- **The Frequentist approach:** Based on the possibility to repeat your experiment.
- **The Bayesian approach:** Based on subjective probability.

Both the methods needs the definition for a **likelihood**.

$$\mathcal{L}(d|h)$$

The Likelihood is a measurement of the “probability” to produce the data you have observed, given your “signal”.

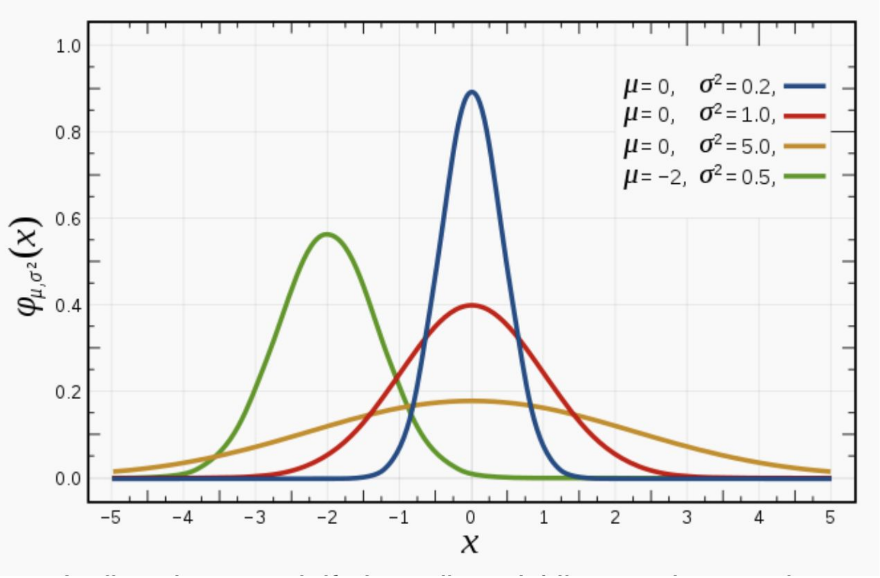
A brief introduction to DA

The most known likelihood, the Gaussian Likelihood.

$$\mathcal{L}(x|\sigma,\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Gaussian likelihood:

- Gives the probability of obtaining a data sample x given two parameters: mean and variance



A brief introduction to DA

Having the likelihood, both the frequentist and the Bayesian method can work.

For frequentists: The likelihood is a probability that can be estimated by repeating my experiment.

Example:

Draw N samples from a Gaussian distribution and then histogram it. From the histogram, you can calculate the mean and the standard deviation.

For Bayesian: The likelihood is a model and the probability is the belief of the experimenter in that model.

Two competing philosophies

Having the likelihood, both the frequentist and the Bayesian method can work.

- **For frequentists**: The likelihood is a probability that can be estimated by repeating my experiment.

  Example:
  Draw N samples from a Gaussian distribution and then histogram it. From the histogram, you can calculate the mean and the standard deviation.

- **For Bayesian**: The likelihood is a model and the probability is the belief of the experimenter in that model.

The frequentist approach

For a frequentist, detecting something means to choose between two hypothesis:

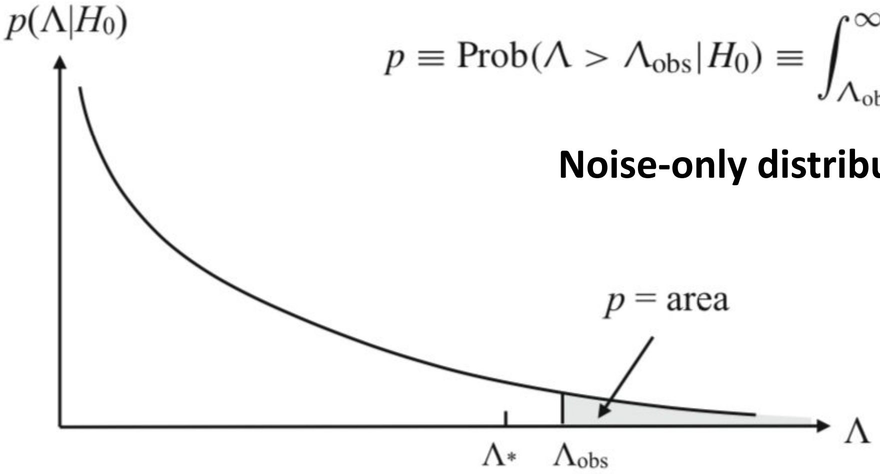
- **Hypothesis 0:** The signal is not present.
- **Hypothesis 1:** The signal is present.

In order to decide what hypothesis we believe, frequentists construct a "detection statistic", i.e. a number for which you can quantify

Λ	Detection statistic. Just a number
$p(\Lambda \vert H_0)$	Distribution of Lambda under the null hypothesis
$p(\Lambda \vert H_1)$	Distribution of Lambda under the hypothesis 1

The frequentist approach: Detection

$$p \equiv \text{Prob}(\Lambda > \Lambda_{\text{obs}} | H_0) \equiv \int_{\Lambda_{\text{obs}}}^{\infty} p(\Lambda|H_0)\, d\Lambda$$



Noise-only distribution of the detection statistic

J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

We reject the Hypothesis 0 using a (1-p)% Confidence Level. If your detection statistic exceed the threshold of the noise-only distribution! Well you have a detection.

False alarm probability: The probability of confusing noise for signal. The noise can be mistaken as a detection!

The frequentist approach

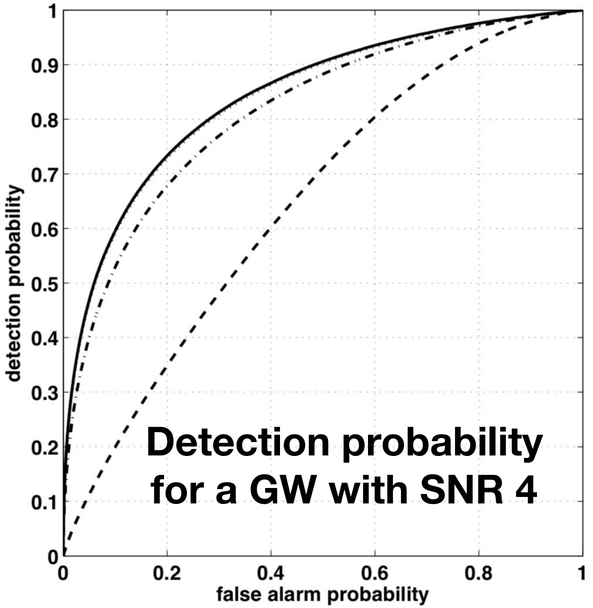
A Good Frequentist method has two variables under control.

$$\alpha \equiv \mathrm{Prob}(\Lambda > \Lambda_* | H_0)$$

$$\beta(a) \equiv \mathrm{Prob}(\Lambda < \Lambda_* | H_a)$$

False alarm probability - The noise can produce triggers

False dismissal probability (1- Detection probability)



When you decide a threshold for your detection problem, you need to find the sweet spot between

- A Low false alarm probability.
- An high detection probability

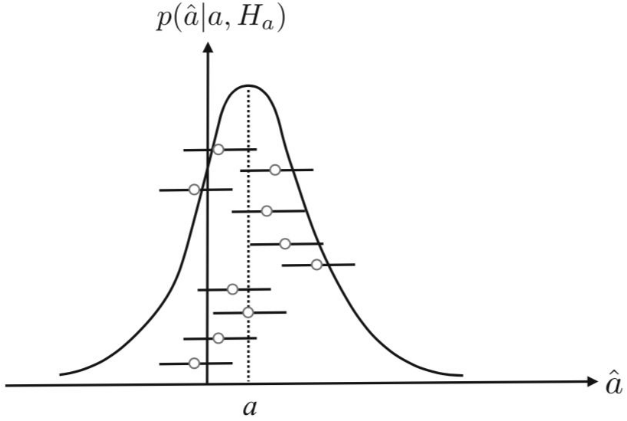
J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

The frequentist approach

Estimating parameters for a frequentist means:

- Build an estimator $\hat{a}$ in the Hypothesis that a signal is present with a parameter a.
- Draw the estimator $\hat{a}$ several times, these will give confidence intervals.
- Repeat this experiment many times.

$$\mathrm{Prob}(a - \Delta < \hat{a} < a + \Delta) = 0.95$$



J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

Bayesian inference

In these notes we are going to touch only the Parameter Estimation for Bayesians.

But before we need to understand the **Bayes Theorem**.

Likelihood

Prior

Posterior

$$p(h|x) = \frac{p(x|h)p(h)}{p(x)}$$

Proof (very easy)

$$p(h|x)p(x) = p(x,h) = p(x|h)p(h)$$

Bayesian inference

In these notes we are going to touch only the Parameter Estimation for Bayesians.

- **Prior**: A belief of the experimenter on the values of the parameters. You can use “flat” priors if you don’t want any information.

- **Likelihood:** We have seen it.

- **Evidence:** The integral of the numerator w.r.t. the parameters. It is a normalisation factor.

- **Posterior:** This is what we want. The probability of the parameters given the data.

Posterior Likelihood Prior

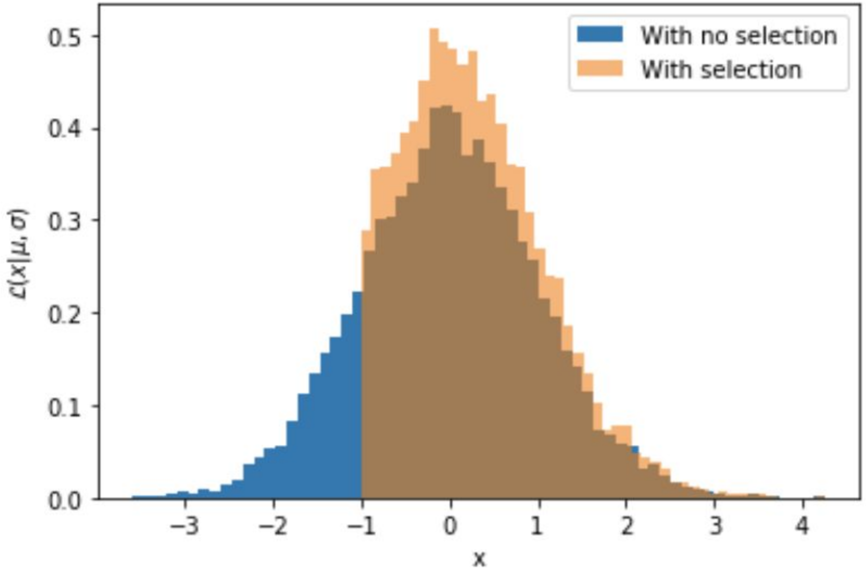
$$p(h|x) = \frac{p(x|h)p(h)}{p(x)}$$

Evidence

Bayesian inference: An example

A random process generating numbers from a normal distribution with mean 0 and variance 1.

Your experiment however gives you only numbers > -1. How you can estimate the mean from these samples?



Bayesian inference: An example

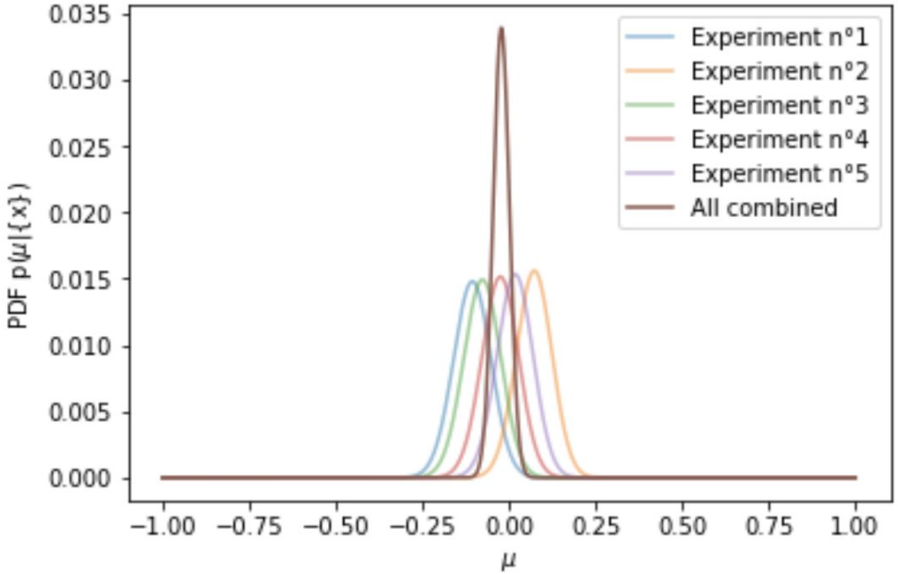
If you include the selection effect, you will converge to the correct value of the mean.

$$p(\mu|\{x\}) \propto \mathcal{L}(\{x\}|\mu)p(\mu) = p(\mu)\prod_i \mathcal{L}(x_i|\mu)$$

You can even do it analytically!

Exercise:

- Show that the expected value of the posterior is the mean.
- Show that the second statistical moment is sigma/srt(N)



Bayesian inference: The selection bias

How to include this?

We should modify the normalisation factor of the likelihood because we know that there are some values of x that we cannot observe.

Likelihood before the selection effect

$$\mathcal{L}(x|\mu) = \frac{\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]}{\int_{-\infty}^{\infty}\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]\mathrm{dx}} = \frac{1}{\sqrt{2\pi\sigma}}\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]$$

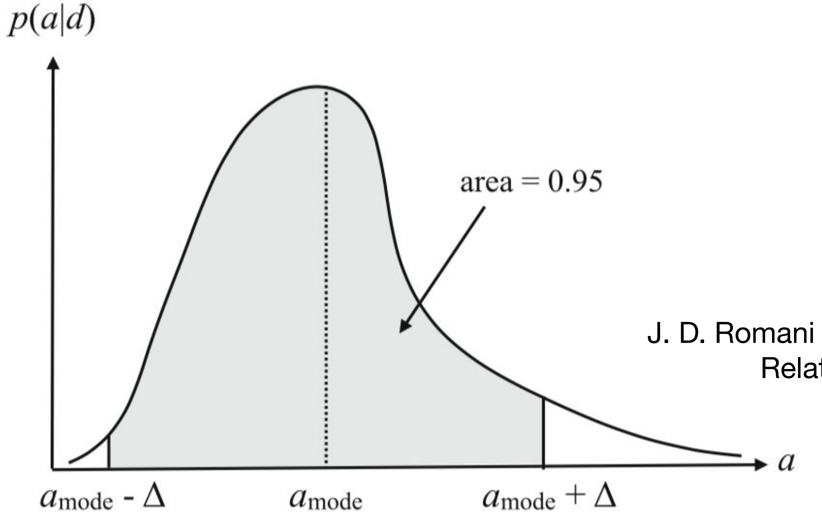
Likelihood after the selection effect

$$\mathcal{L}(x|\mu) = \frac{\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]}{\int_{x_{thr}}^{\infty}\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]\mathrm{dx}} = \frac{\exp[-(\mathrm{x}-\mu)^2/(2\sigma^2)]}{I(\mu, x_{thr})}$$

Bayesian inference: Parameters

The process of a Bayesian Learner

$$p(a|x) = \frac{p(x|a)p(a)}{\int p(x|a)p(a)}$$



J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

The posterior is a distribution

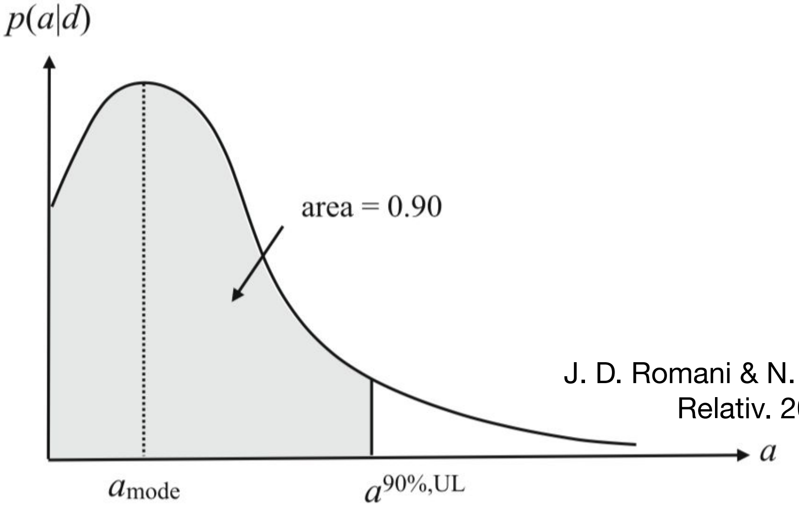
When the posterior goes to zero

You can provide Constraints

Bayesian inference: Parameters

The process of a Bayesian Learner

$$p(a|x) = \frac{p(x|a)p(a)}{\int p(x|a)p(a)}$$



J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

The posterior is a distribution

When the posterior does not go to zero

You can provide upper limits

The gaussian example

Click on this link for more

Part 1 - The end

INFN
Istituto Nazionale di Fisica Nucleare

GravitySirens

Funded by the European Union

erc
European Research Council
Established by the European Commission