Data analysis techniques for gravitational waves

2nd Manitou School July 7th 2023

Dr. S. Mastrogiovanni

(mastrosi@roma1.infn.it)

Outline of the lectures

- **What to expect:** A soft introduction to many topics that will make you able to understand what LIGO, Virgo, KAGRA are publishing (from a personal angle).
    - Difference between Bayesian and frequentist statistics.
    - Searching for GW signals in time and frequency domain.
    - Bayesian parameter estimation
    - Population inference.

Some references

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

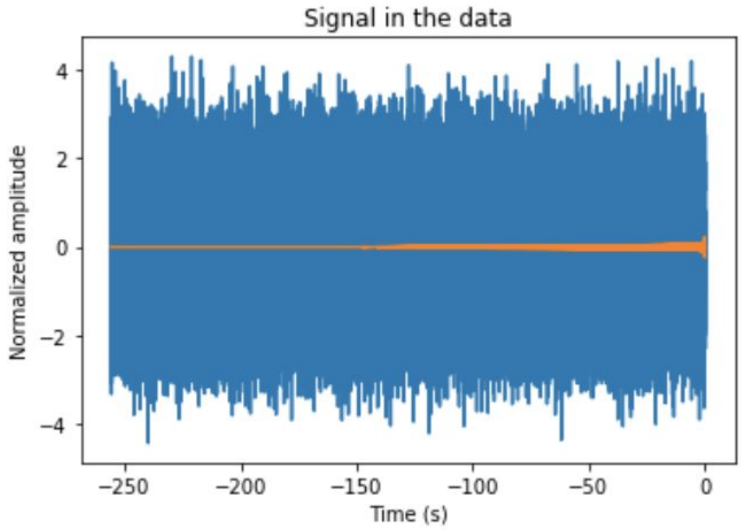
- A good introduction to GW DA aspects (from the GW stochastic searches)

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

Introduction to Data Analysis

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 Data Analysis

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”.

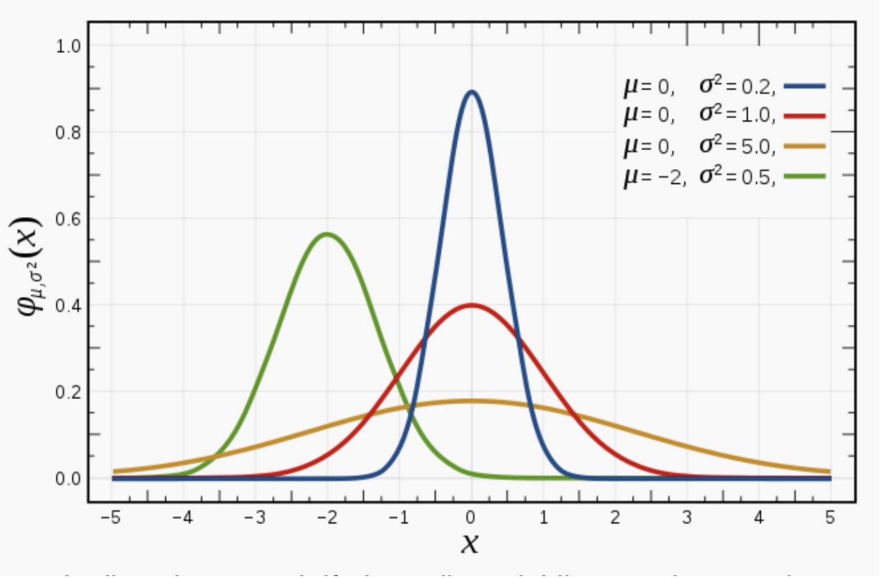
Introduction to Data Analysis

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



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.

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.

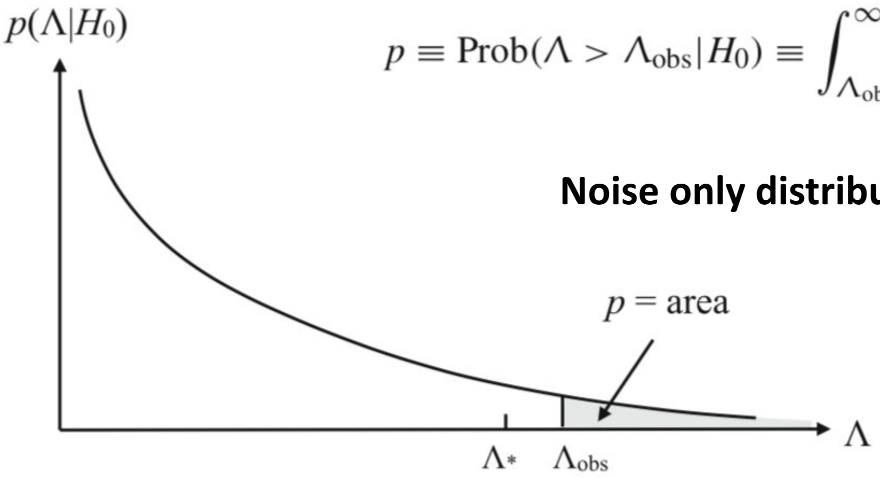
For a frequentist, detecting something means to choose between two hypothesis:

- **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
$p(\Lambda \| H_0)$	Distribution of Lambda under the null hypothesis
$p(\Lambda \| H_1)$	Distribution of Lambda under the hypothesis 1

Introduction to Data Analysis: Detection problem (freq.)



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

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.

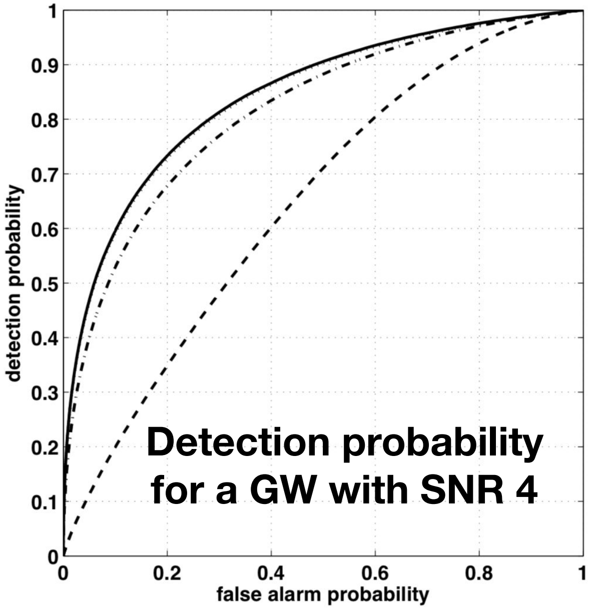
False alarm probability: The probability of confusing noise for signal. The noise can be mistaken as a detection!

Introduction to Data Analysis: Detection problem (freq.)

A Good Frequentist method has two variables under control.

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

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



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

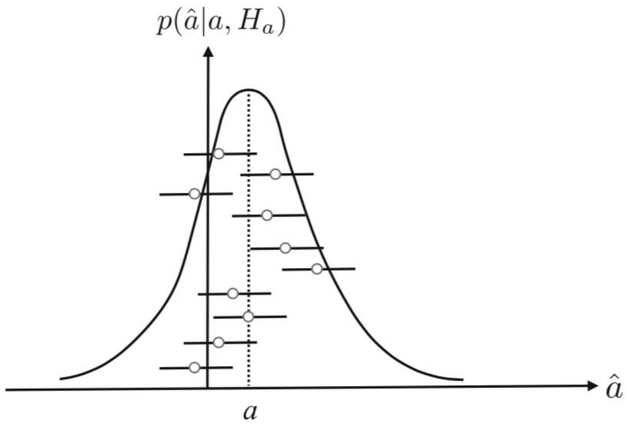
J. D. Romani & N. Cornish, Living Rev. Relativ. 2017, 20:2

Introduction to Data Analysis: Parameter estimation (freq.)

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

Introduction to 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)

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

Introduction to 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.

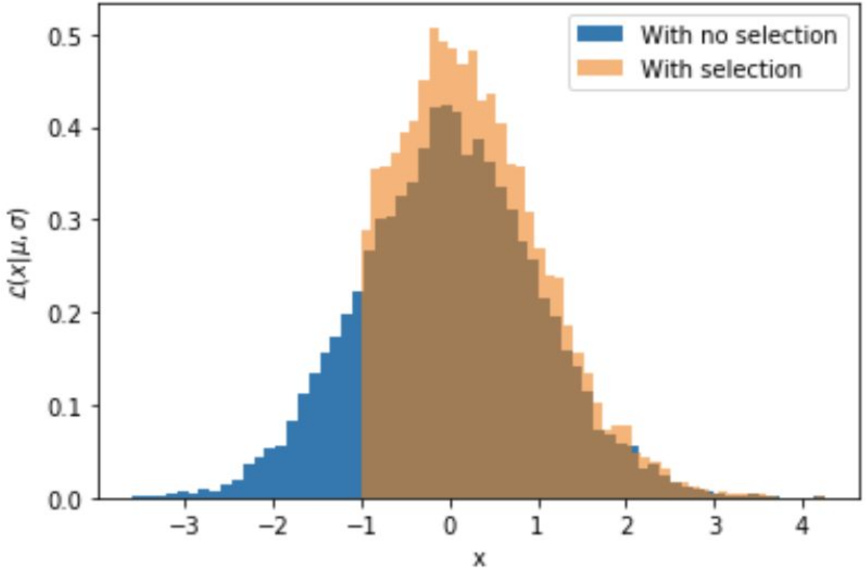
Posterior · Likelihood · Prior · Evidence

$$\underbrace{p(h|x)}_{\text{Posterior}} = \frac{\overbrace{p(x|h)}^{\text{Likelihood}}\overbrace{p(h)}^{\text{Prior}}}{\underbrace{p(x)}_{\text{Evidence}}}$$

Introduction to Bayesian inference

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

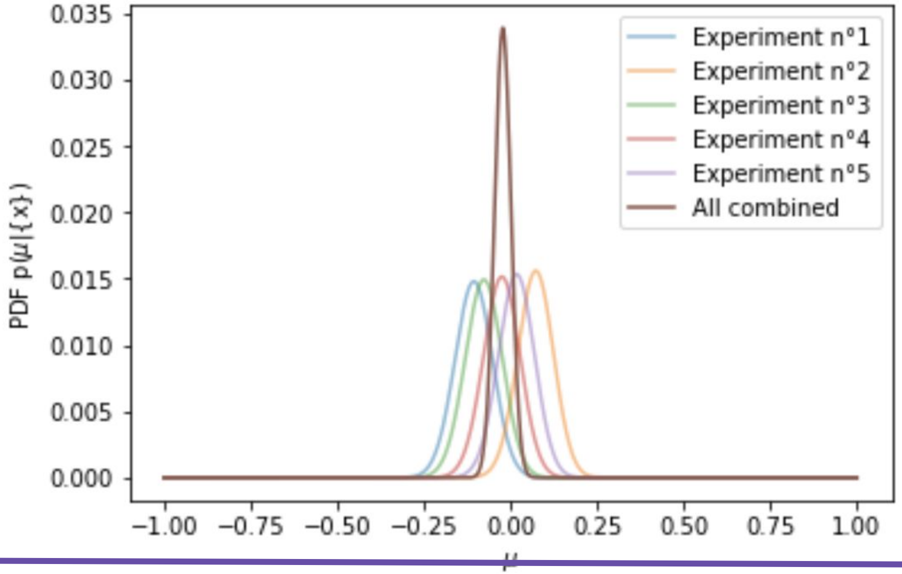
Your experiment however gives you only numbers > -1. How you can estimate the mean from these samples?



Introduction to Bayesian inference

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

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



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

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

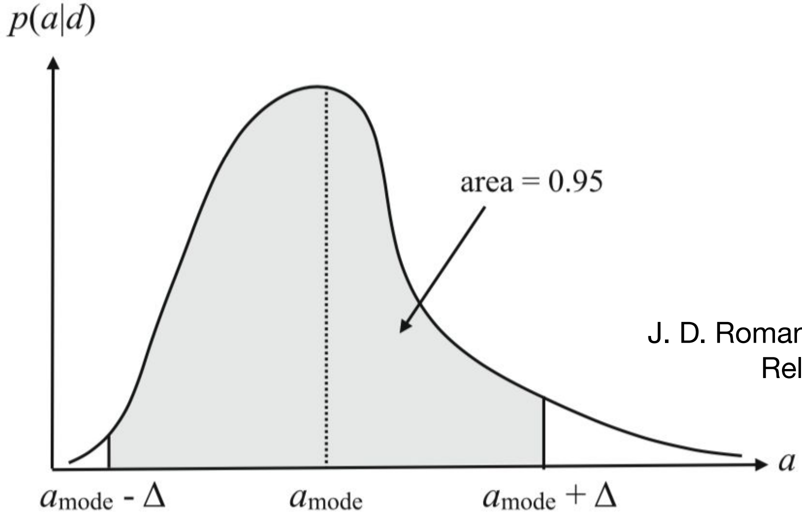
Likelihood after the selection effect

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

Introduction to Bayesian inference: Parameter estimation

The process of a Bayesian Lerner

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



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

The posterior is a distribution

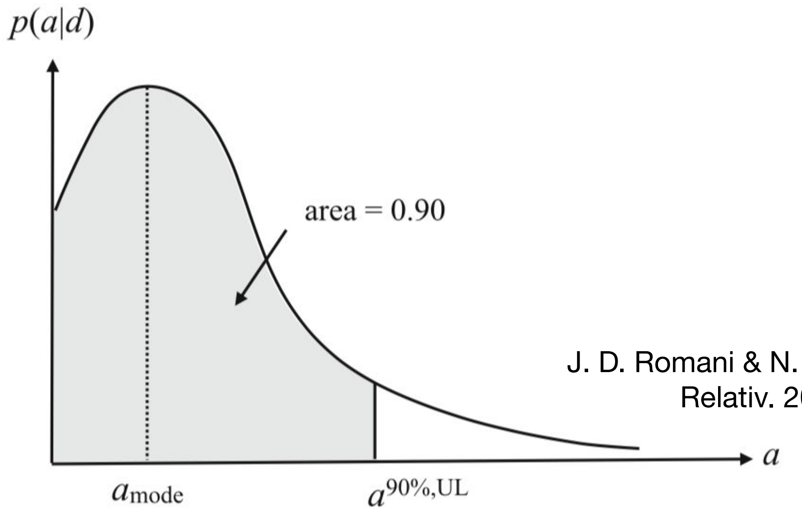
When the posterior goes to zero

You can provide Constraints

Introduction to Bayesian inference: Parameter estimation

The process of a Bayesian Lerner

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



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

Click on this link for more

Introduction to GW spectral analysis

Introduction to time series

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.

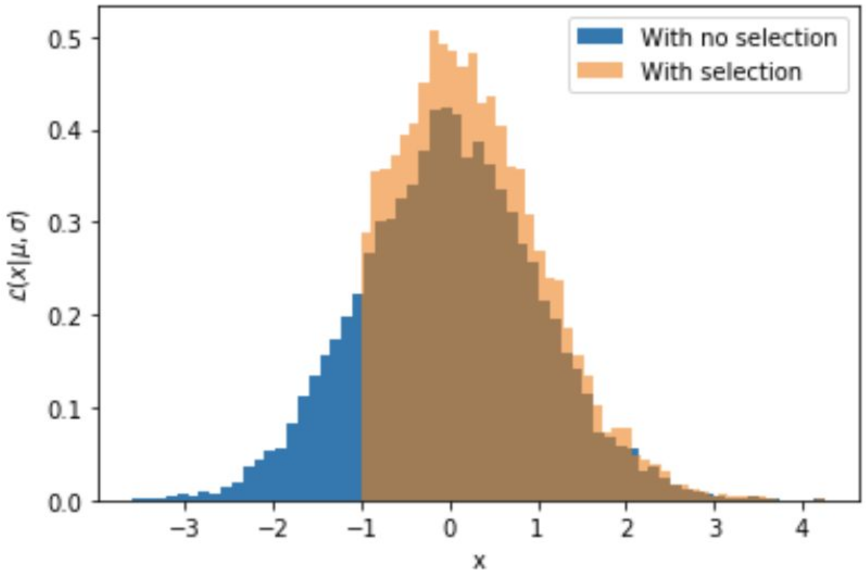
$$d = \{n_1 + F[h_1(\theta)], \ldots, n_N + F[h_N(\theta)]\}$$

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

Introduction to Bayesian inference

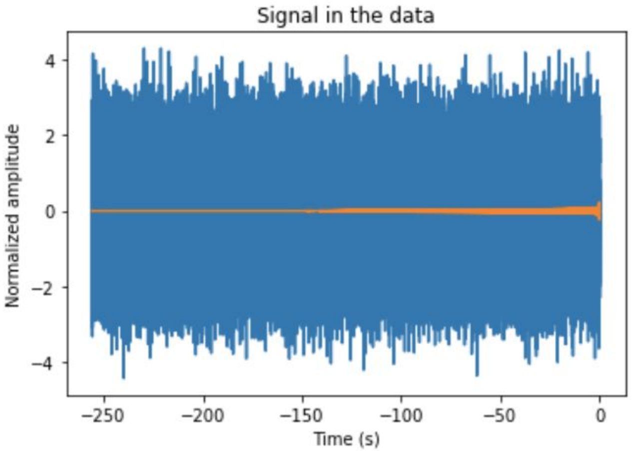
A random process generating numbers from a normal distribution with mean 0 and variance 1.

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



Time series

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



How do we visualise that a signal is present in here?

We work in the frequency domain

The Discrete Fourier Transform

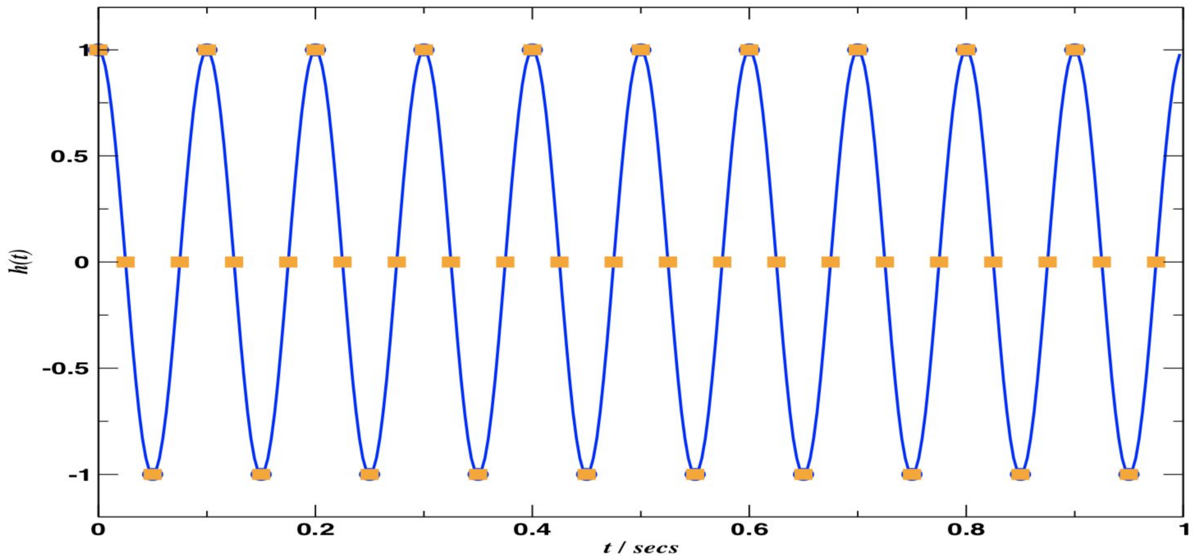
The Fourier transform (FFT for fast) is a procedure which allows you to compute the frequency components of your **time series**.

Change of basis: From the time domain to the frequency domain.

$$x(f_i) = \sum_i x(t_i) e^{-2\pi i f_i t_i} dt$$

The Discrete Fourier Transform

The Fourier transform (FFT for fast) is a procedure which allows you to compute the frequency components of your **time series**.

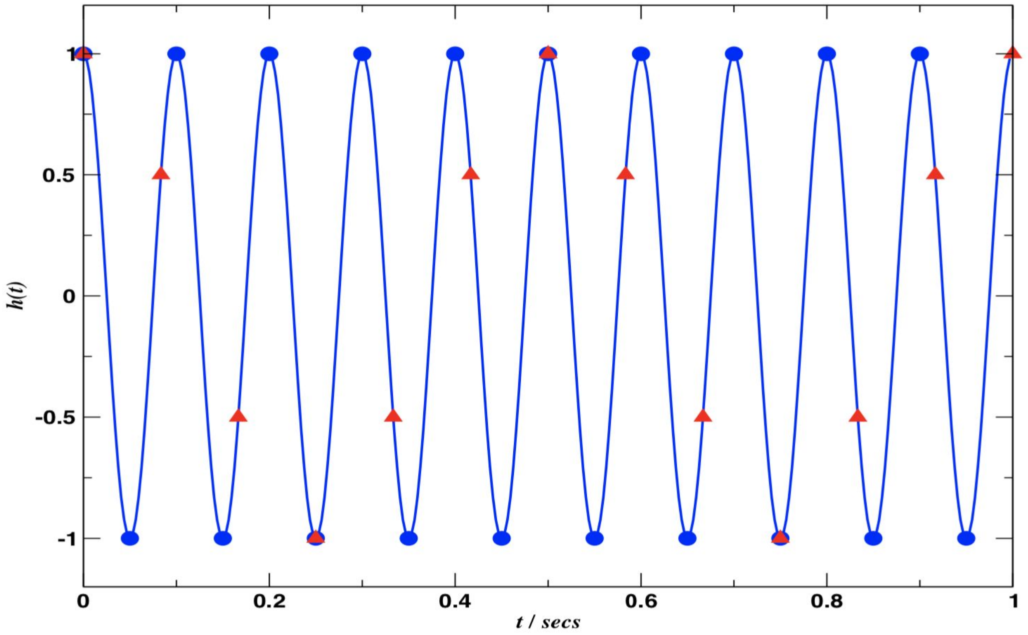


Is it oversampled?

$f_s = 4f$

The Discrete Fourier Transform

When you are working with discrete time series, you have some fundamental properties

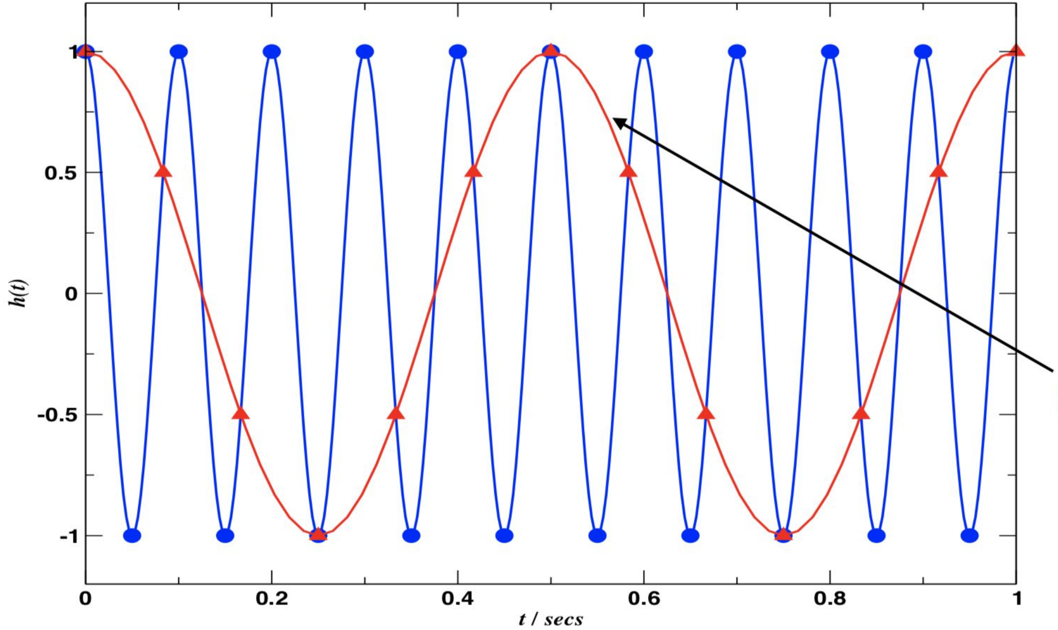


Looks better but!

$f_s = 1.2f$

The Discrete Fourier Transform

When you are working with discrete time series, you have some fundamental properties



$f_s = 1.2f$

Corresponds to a wave at 2 Hz

The Discrete Fourier Transform

When you are working with discrete time series, you have some fundamental properties

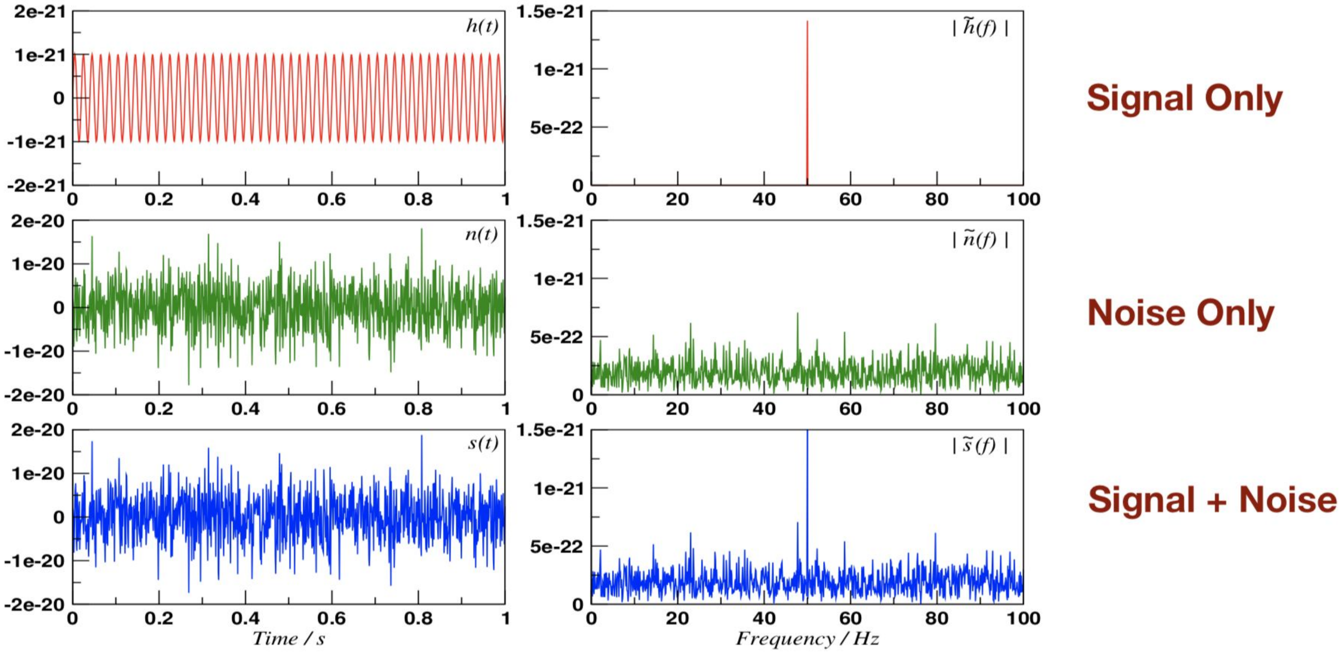
- Fourier transform resolution: $$\delta f = T_{\rm obs}^{-1}$$

- Fourier transform span: You can calculate frequency components up to half of the sampling rate (Nyquist frequency)

$$f_{\rm max} = 0.5 \delta t^{-1}$$

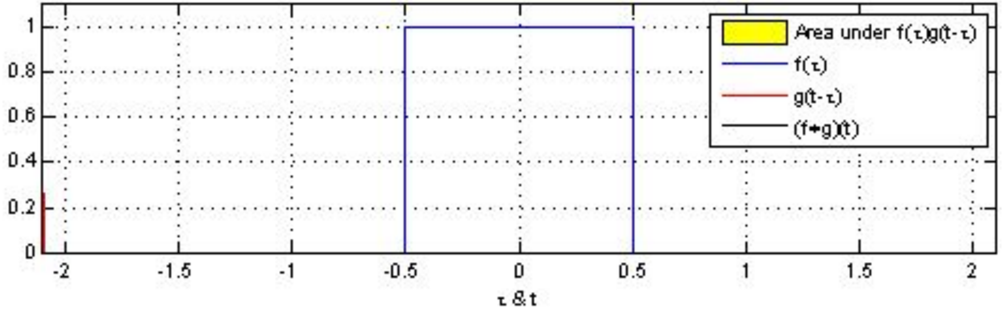
Signals in frequency

If you FT the entire time series, then you can compute the spectra of the signal



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



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

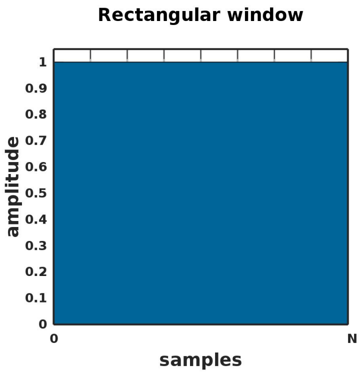
The convolution theorem

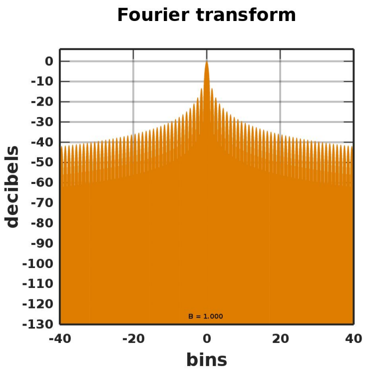
The convolution theorem relates multiplication of functions in time/frequency domain to convolutions in frequency/time domain

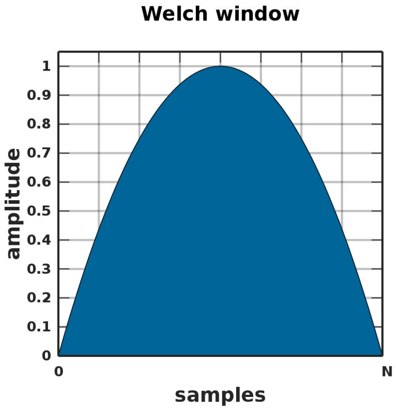
$$\mathcal{F}[\int_t f(t-\tau)g(t)dt] = \tilde{f}(\omega)\tilde{g}(\omega)$$

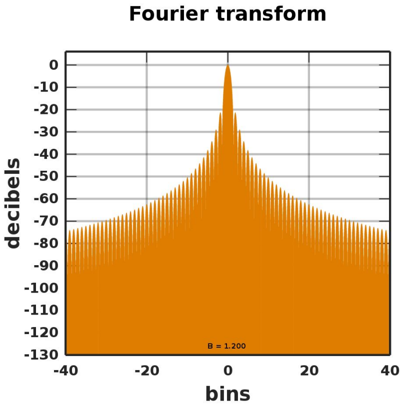
$$\mathcal{F}^{-1}[\int_\omega \tilde{f}(\omega-\omega')\tilde{g}(\omega)d\omega] = f(t)g(t)$$

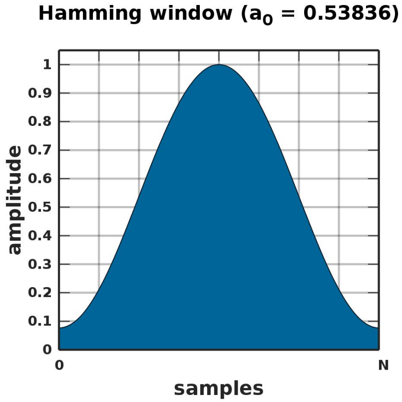
The convolution theorem and windows

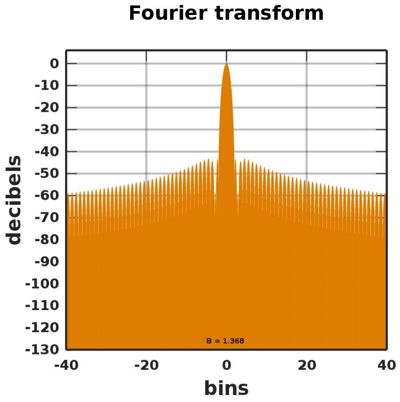


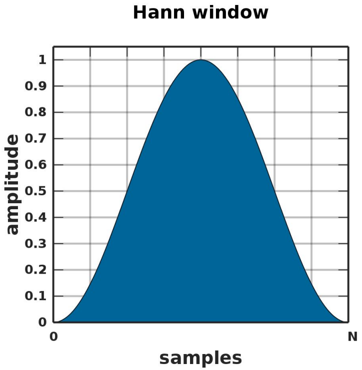


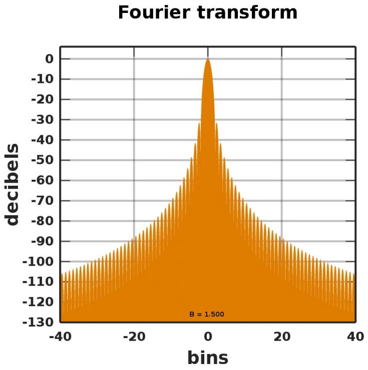












Signals in frequency

Spectra are good for visualising signals which are:

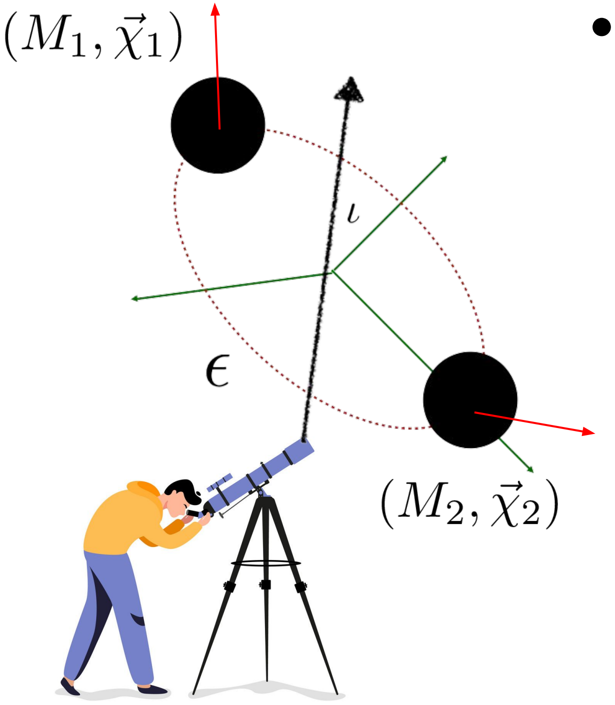
- Continuous signals: They last in all your time series
- Deeply buried into the noise: The high resolution of FT helps.
- Almost-monochromatic: They are sine waves for combination of them.

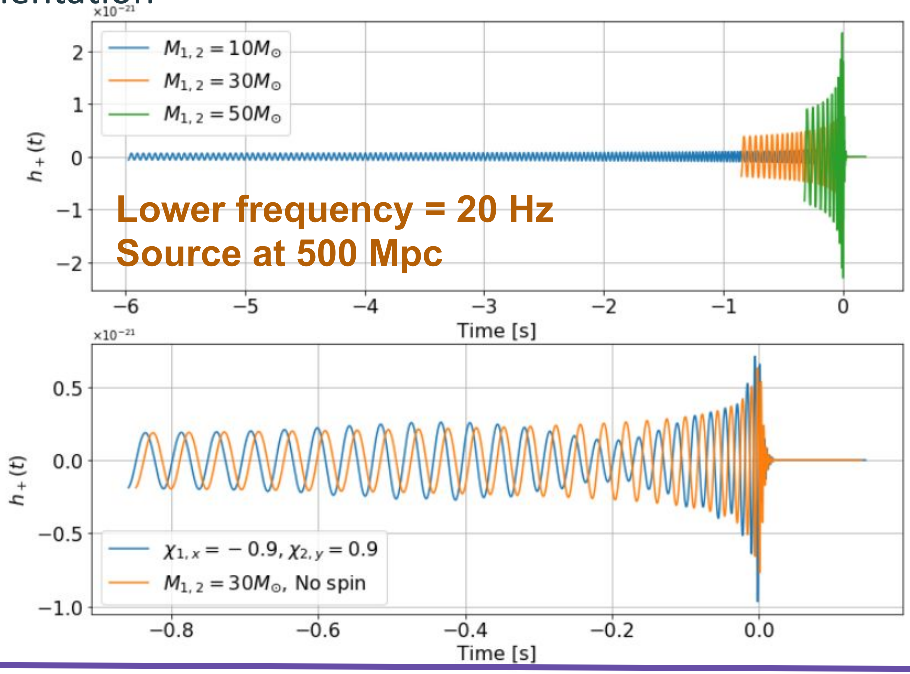
If you want instead to visualise "transient signals", that changes frequency very fast, you need to FT little chunks of data, in which the frequency is more or less constant (w.r.t. the resolution)

The CBC signal

The parameters:

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

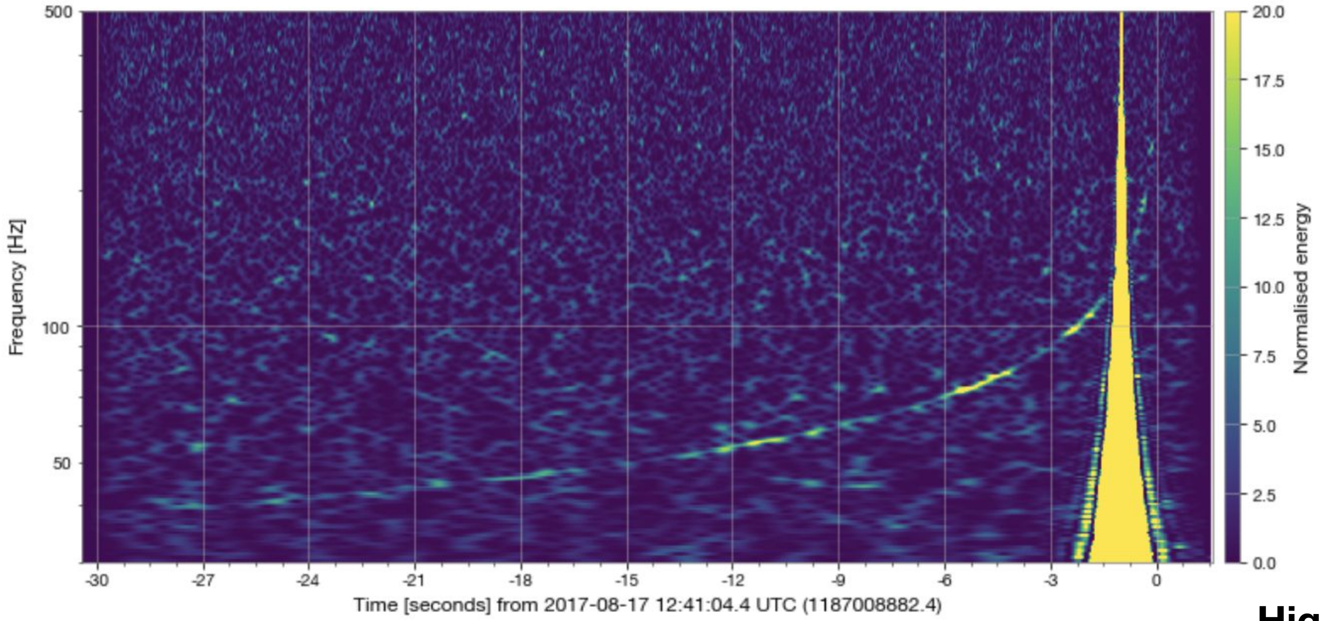




The CBC signal: Time-frequency maps

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

Higher time resolution



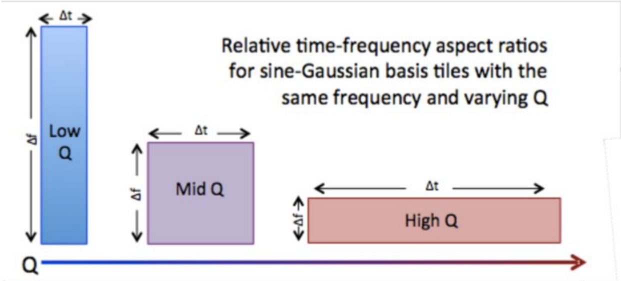
Higher frequency resolution

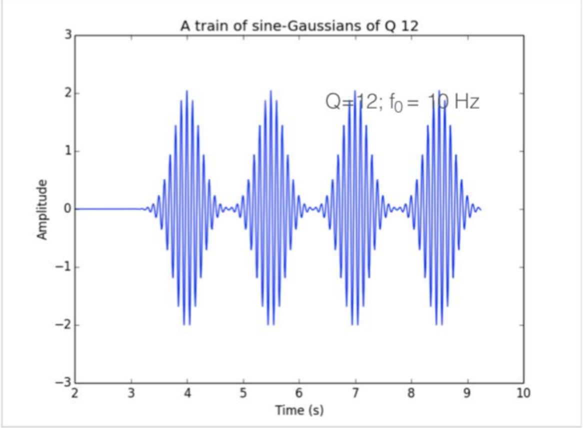
Time-frequency maps: Q-transform

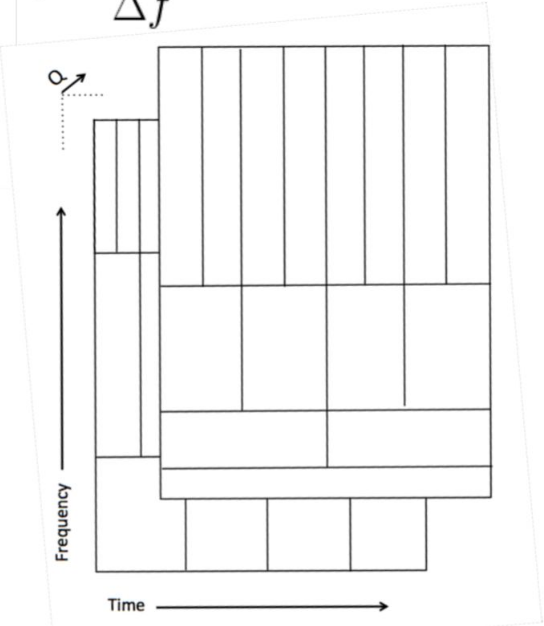
Time-Frequency plots for LIGO are generated with a special FT.

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

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







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

Link

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 quintity(SNR)

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

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

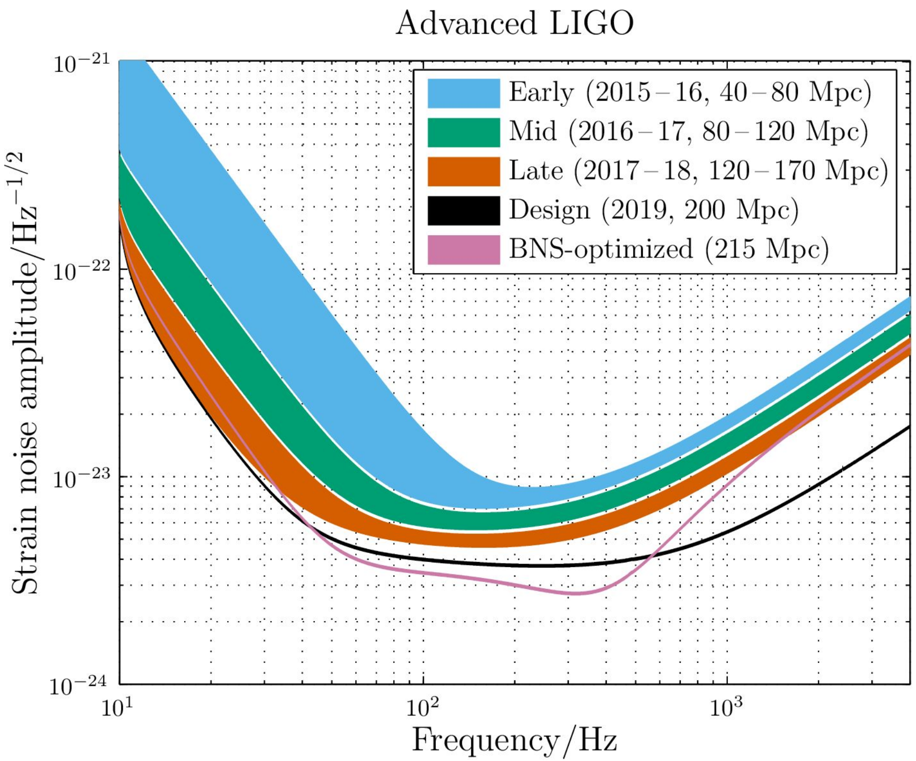
The noise background that you have seen in the previous plots it is called power spectral density (PSD)

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

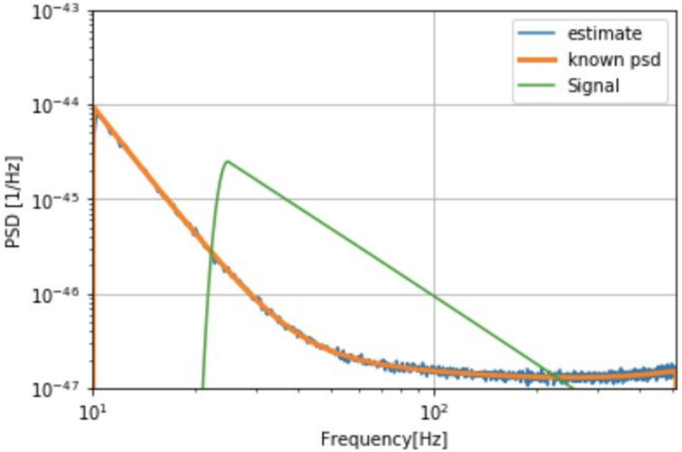
$$F(f) = h(f)e^{-2\pi i f\tau}$$

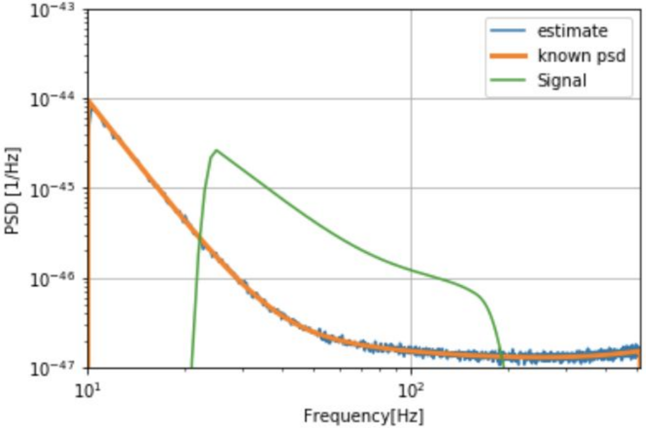
Introduction to matched filtering

With these choice of filter, the SNR become

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

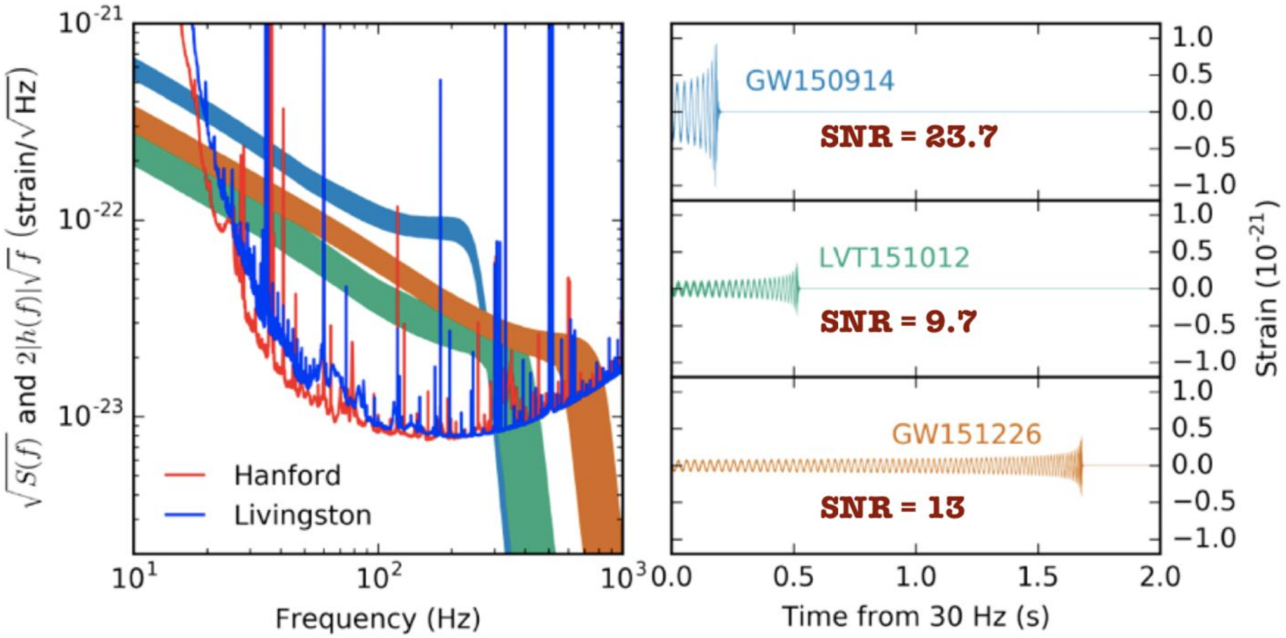
The SNR is basically an area!





Introduction to matched filtering

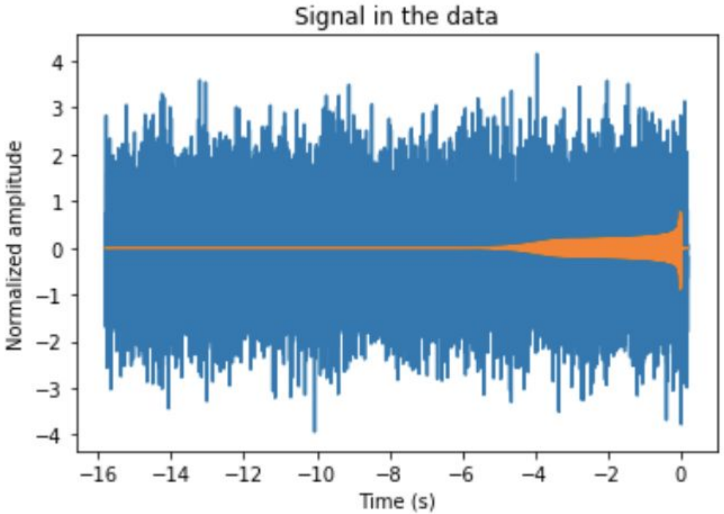
Signal to noise events of several events at comparison

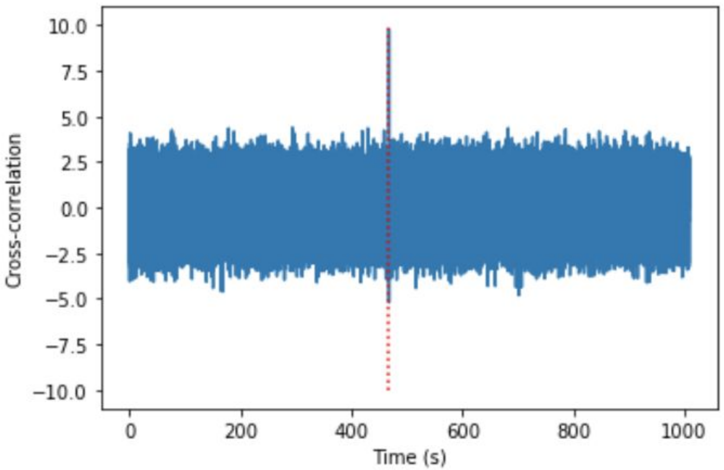


Introduction to matched filtering

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.



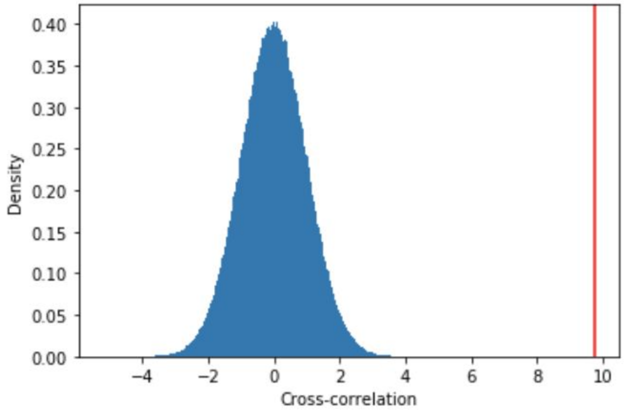


Introduction to matched filtering

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

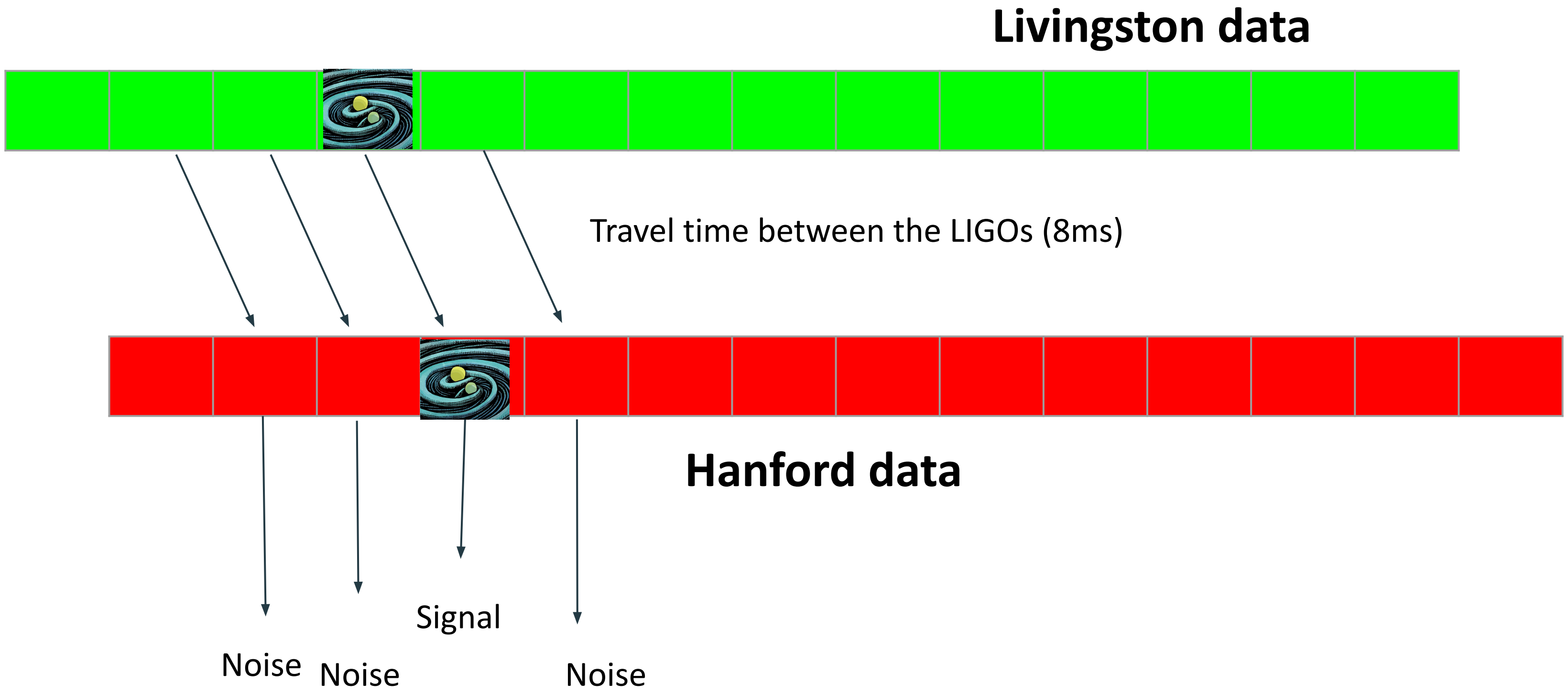
2. When we match the signal, we obtain a very strong outlier, which is not compatible with the background distribution of the noise.

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



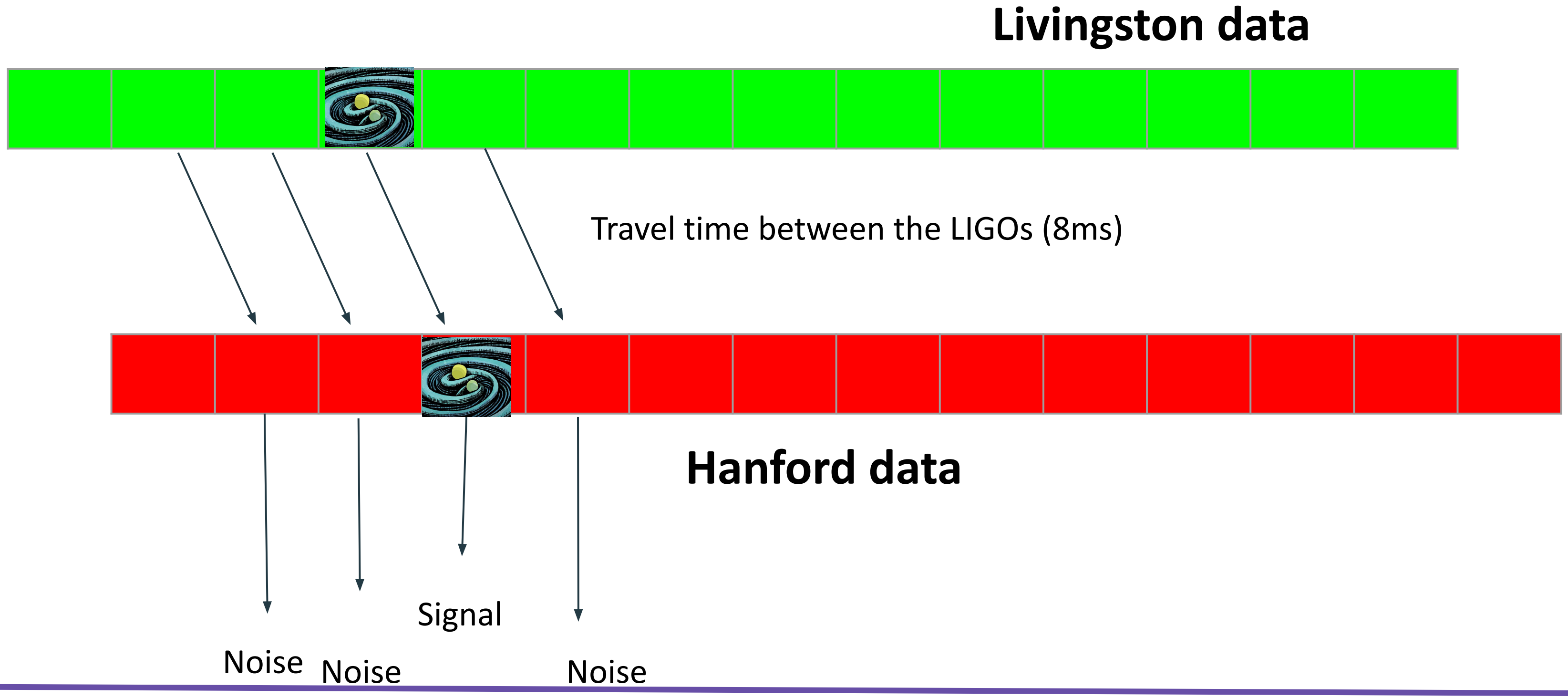
Introduction to matched filtering

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.



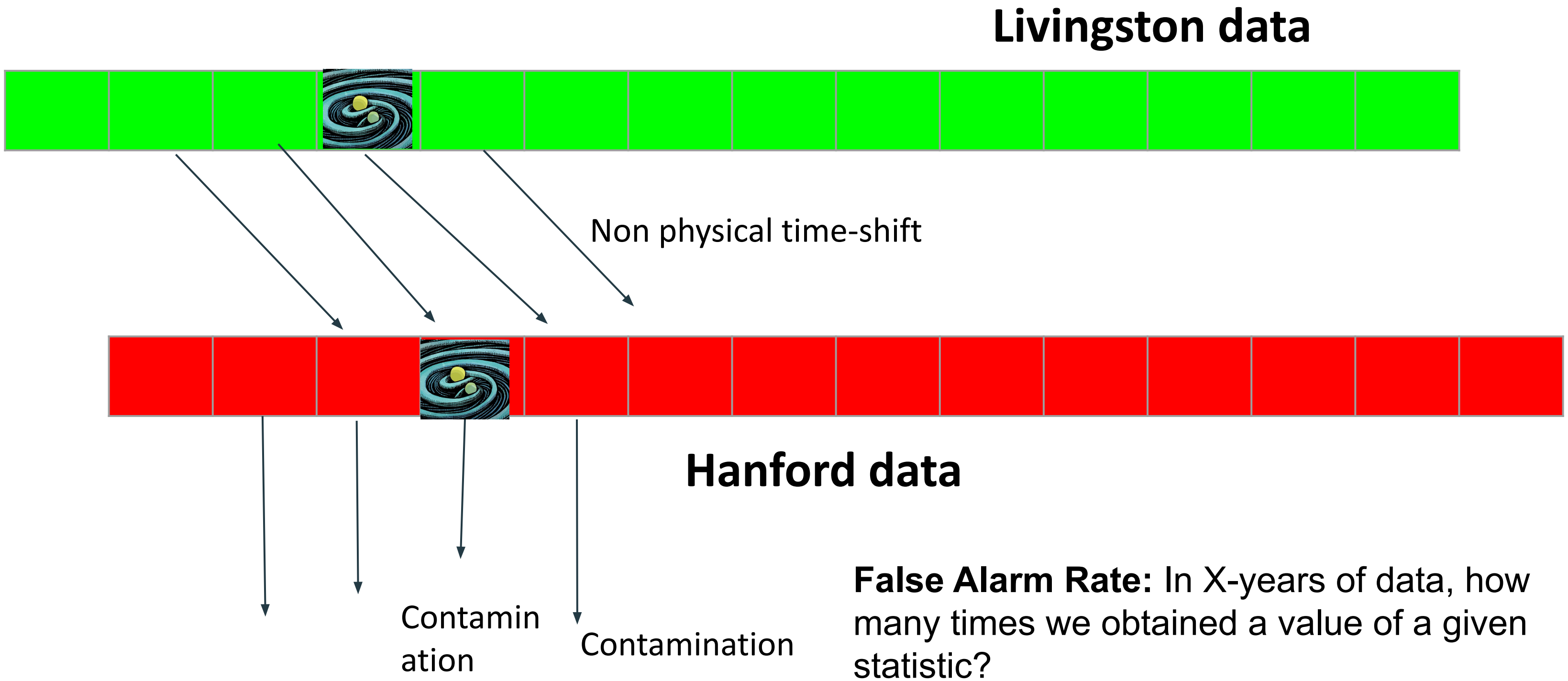
Introduction to matched filtering

How do we generate noise backgrounds? In a real search we never know when the signal is present or not...



Introduction to matched filtering

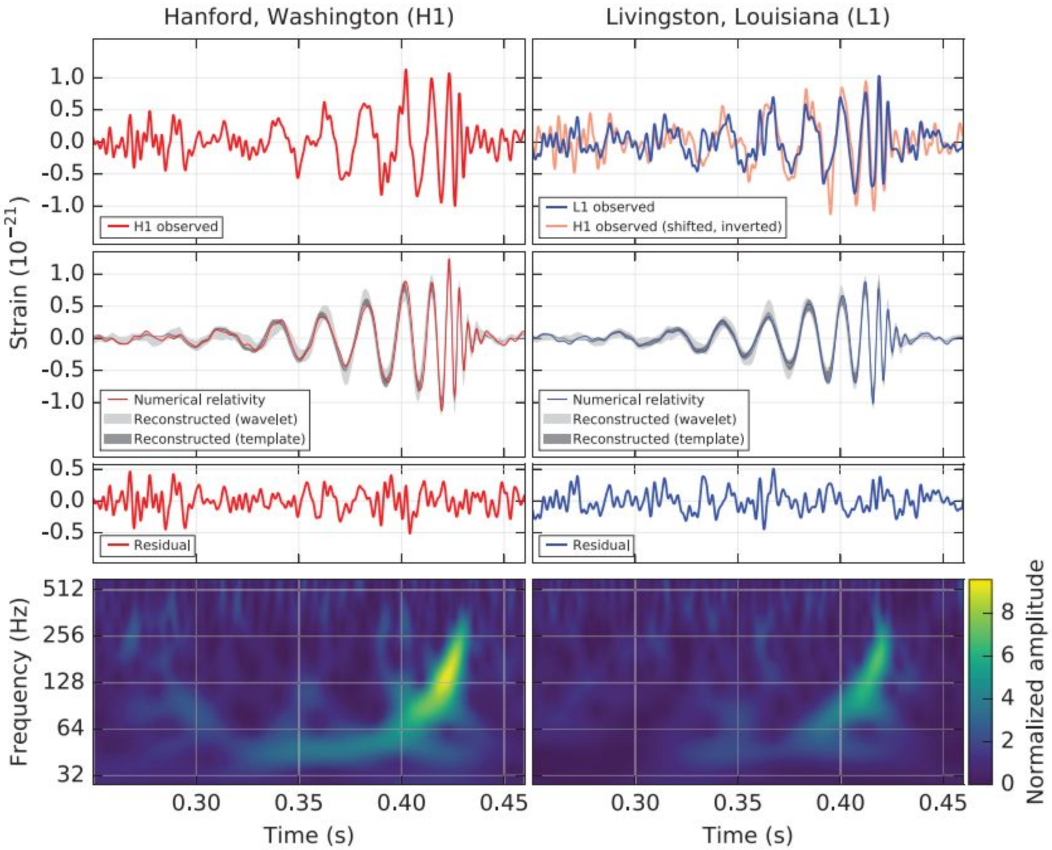
To generate noise backgrounds we perform the same search but matching the data between interferometers ``*wrongly*``



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

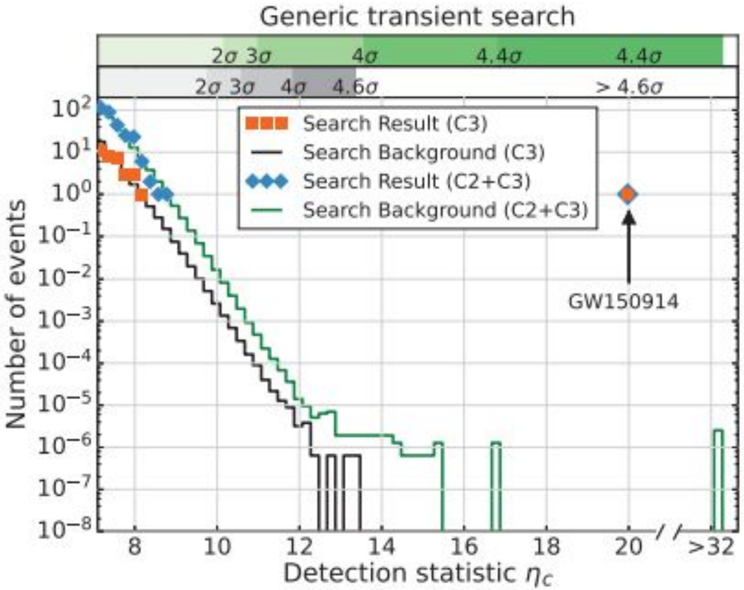
Introduction to matched filtering

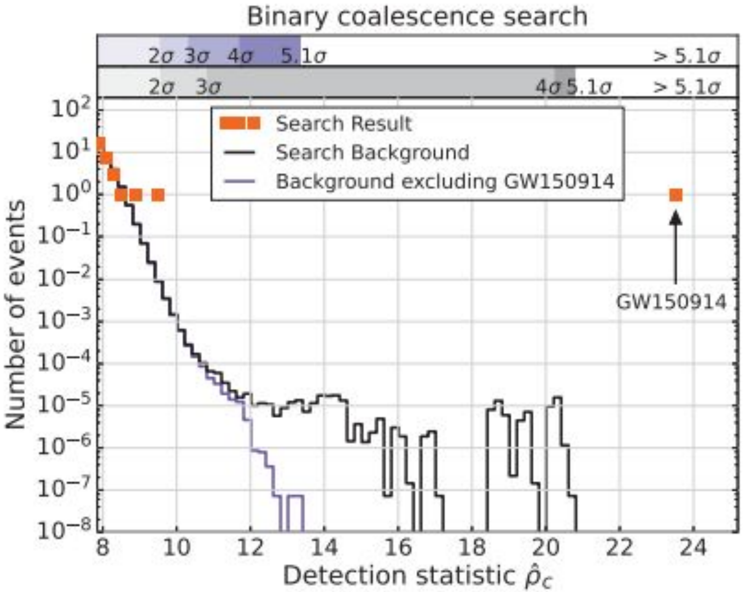
- GW150914 was detected on September 14, 2015 at 09:50:45 UTC.
- The signal was observed between 35 and 250. The strain peak was around 1e-21.
- The signal had a SNR of 24.



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.





Matched filtering in action (from GW open data workshop)

[Link]

Introduction to Bayesian parameter estimation

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 Bayesian parameter estimation

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

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

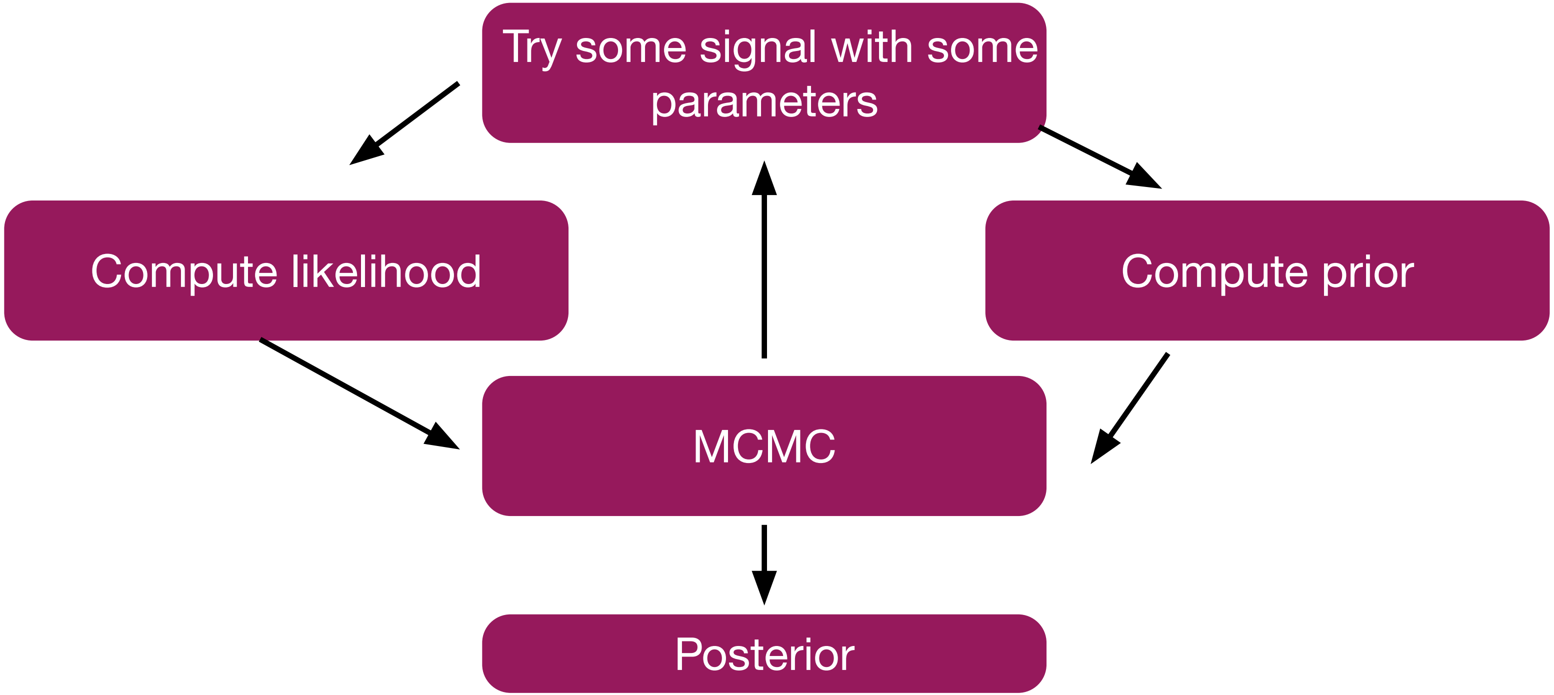
The scalar product that we have introduced, is directly related to the definition of the SNR. In fact, as we have seen before.

The higher the SNR, the more peaked is the likelihood

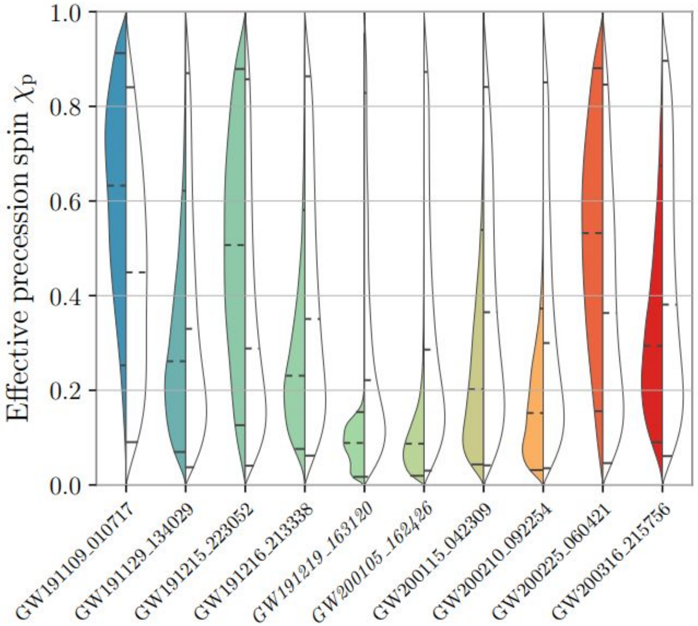
$$SNR^2 \equiv < h|h > = \int_{f_l}^{f_u} \frac{|h(f)|^2}{S_n(f)} df$$

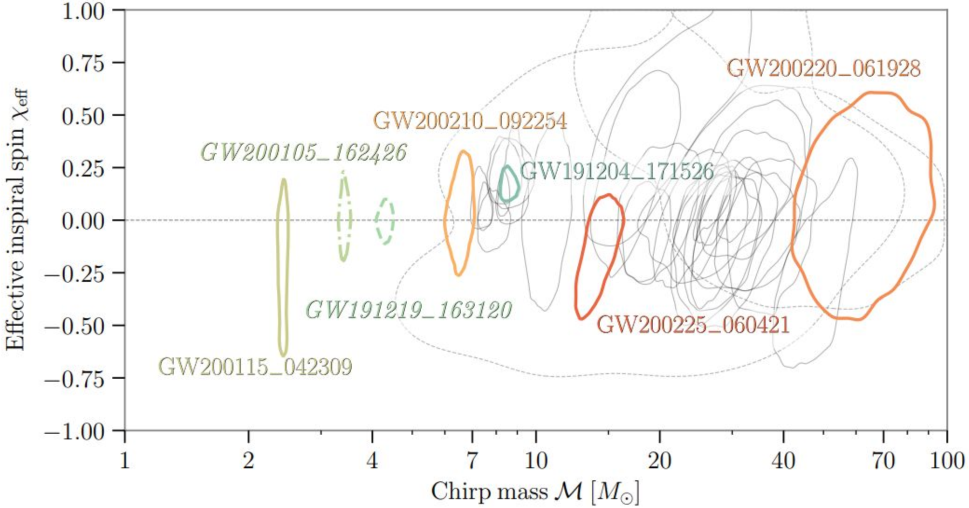
Introduction to Bayesian parameter estimation

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



Introduction to Bayesian parameter estimation





Introduction to Hierarchical Bayesian inference

Physics		Statistics
Particles energy **E** are generated with a gaussian distribution	⟺	$p_{\rm pop}(E\vert\Lambda) = \frac{1}{N}\frac{dN}{dE}$
Particles arrive at our detectors, due to noise I measure $\mathbf{E_{obs}}$	⟺	$\mathcal{L}_n(E_{\rm obs}\vert E)$
My experiment can detect $\mathbf{E_{obs}}$ only above a certain threshold	⟺	$p_{\rm det}(E) = \int_{E_{\rm thr}}^{\infty} \mathcal{L}_n(E_{\rm obs}\vert E) dE_{\rm obs}$

Introduction to Hierarchical Bayesian inference

The likelihood for an inhomogeneous Poisson process in presence of selection biases, for a **constant rate in detector time**, is (see Mandel+ 2018 MNRAS, Vitale+ 2020)

Rate

$$\mathcal{L}(x|\Lambda) \propto e^{-N_{\rm exp}} \prod_{i=1}^{N_{\rm obs}} T_{\rm obs} \int \mathcal{L}_n(x|\theta,\Lambda) \frac{dN}{dtd\theta} d\theta$$

Noise process

Expected number of detections

$$N_{\rm exp} = T_{\rm obs} \int p_{\rm det}(\theta,\Lambda) \frac{dN}{dtd\theta} d\theta.$$

Introduction to Hierarchical Bayesian inference

If we decide to use a 1/Nexp prior, the hierarchical likelihood can be reduced to the form below

$$\mathcal{L}(x|\Lambda) \propto \prod_{i=1}^{N_{\rm obs}} \frac{\int \mathcal{L}_n(x|\theta,\Lambda)\frac{dN}{dt d\theta}d\theta}{\int p_{\rm det}(\theta,\Lambda)\frac{dN}{dt d\theta}d\theta}$$

Scale-free likelihood

Introduction to Hierarchical Bayesian inference

From the previous slides we have seen that to calculate the likelihood we need to calculate two integrals, the integral over the parameters of the detected events and the integral to calculate Nexp.

Integral over the events (numerator): Done summing over posterior samples

$$\int \mathcal{L}_n(x|\theta,\Lambda)\frac{dN}{dtd\theta}d\theta \approx \frac{1}{N_s}\sum_j^{N_s}\frac{1}{\pi_{\rm PE}(\theta_j)}\frac{dN}{dtd\theta}\Big|_j$$

Number of samples

Prior used for Posterior samples

Introduction to Hierarchical Bayesian inference

From the previous slides we have seen that to calculate the likelihood we need to calculate two integrals, the integral over the parameters of the detected events and the integral to calculate Nexp.

Integral over the events (numerator): Done summing over posterior samples

Number of detected events

$$N_{\rm exp} \approx \frac{T_{\rm obs}}{N_{\rm gen}} \sum_{j}^{N_{\rm det}} \frac{1}{\pi_{\rm inj}(\theta_j)} \frac{dN}{dt d\theta}\Big|_j$$

Number of injections generated (even not detected)

Prior used for injections

Introduction to Hierarchical Bayesian inference

The hierarchical likelihood computed numerically is given by the equation below

$$\mathcal{L}(x|\Lambda) \approx e^{-N_{\rm exp}} \prod_{i=1}^{N_{\rm obs}} \frac{T_{\rm obs}}{N_s} \sum_{j=1}^{N_s} \frac{1}{\pi(\theta_{j,i})} \frac{dN}{dt d\theta}\bigg|_{j,i}$$

Inferring the rate of a gaussian process

Link