GWD Data Analysis 1

Detection and Parameter Estimation for Compact Binary Coalescences

Qian Hu
Qian.Hu@glasgow.ac.uk

Where are we?

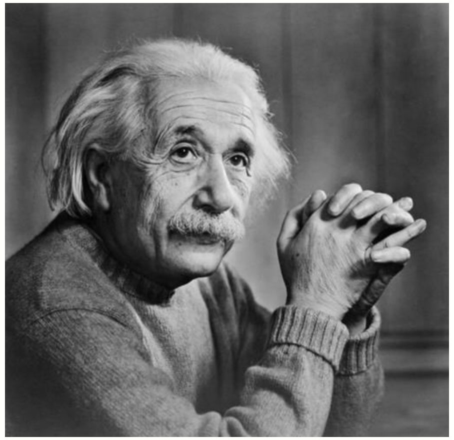
University of Glasgow | IGR

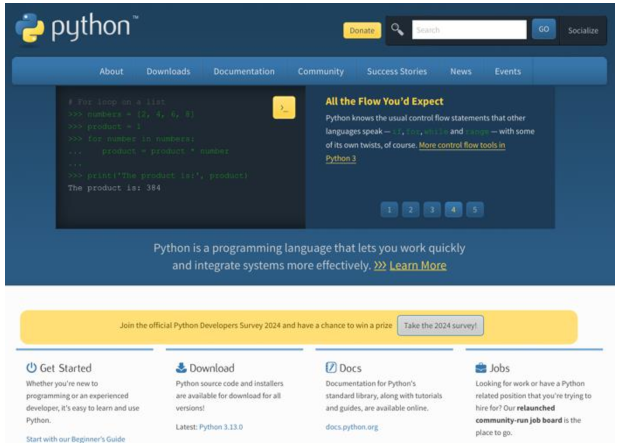
Gravitational Wave Astronomy

Theory → Experiment (Observation) → Data Analysis

Data Analysis → Astrophysics, Cosmology, Gravity theories, Dense stars, black holes, Multi-messenger, Other new physics…

…………………………We are here!

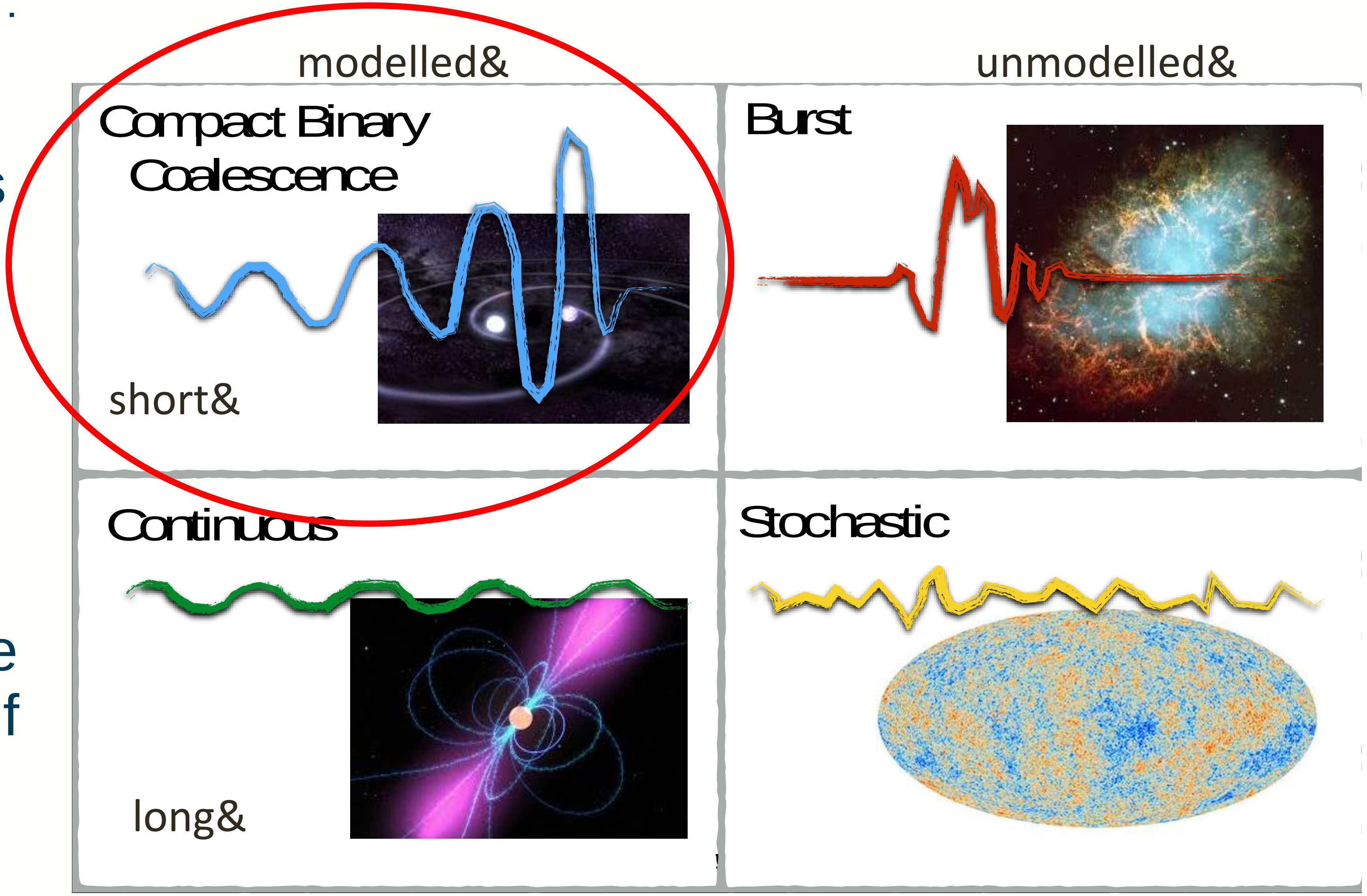




Course Goals

We are trying to understand…

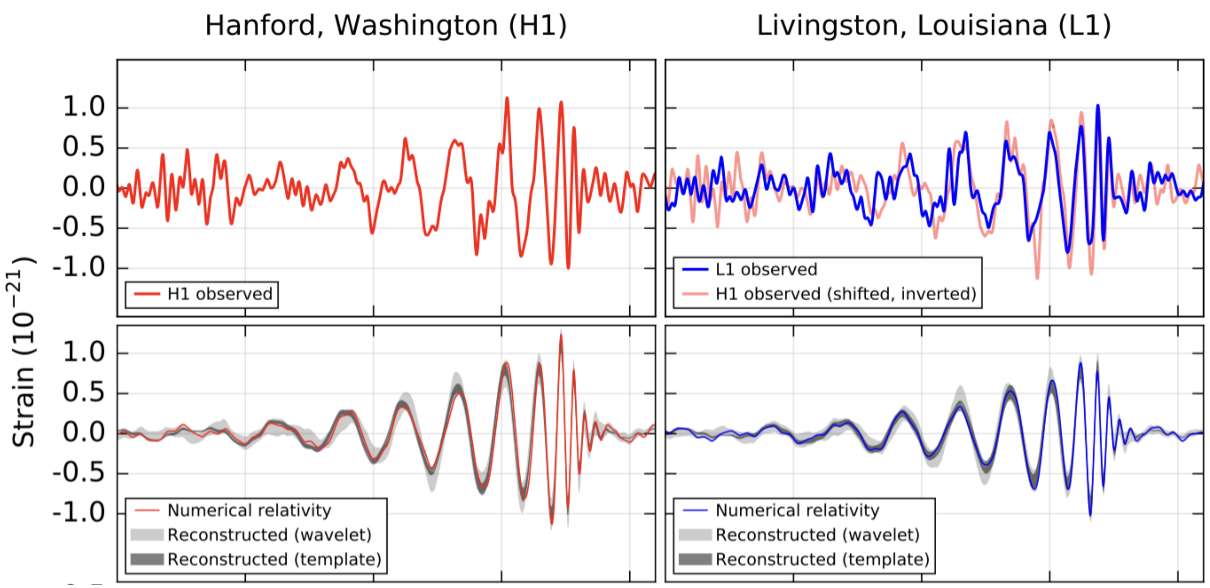
- The statistical behaviors of noises in GW detectors
- How to detect CBC signals
- How to estimate the source properties of CBC signals

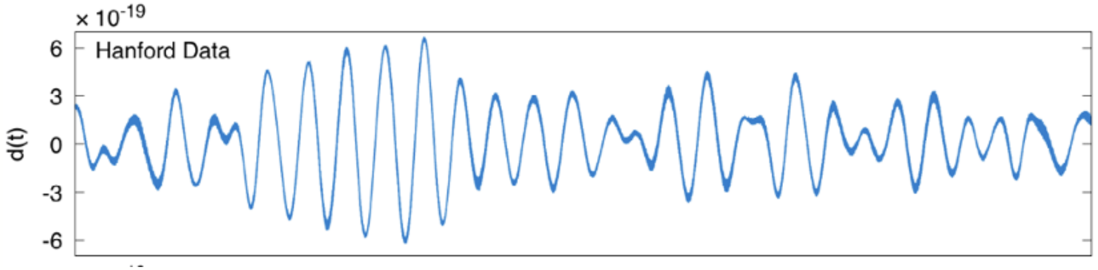


Understanding the noise

The data looks like…

The first detection: GW150914





What the public see

arXiv:1602.03837

Raw data

arXiv:1908.11170

Noise

Generic description

- GW detector noise is a series of **unknown fluctuations** *n* in the detector output *d* caused by the various noise sources.
- It is added linearly to GW signal *h:* $d(t) = h(t) + n(t)$, or in frequency domain, $\tilde{d}(f) = \tilde{h}(f) + \tilde{n}(f)$

$$\tilde{d}(f) = \int_{-\infty}^{+\infty} d(t) e^{-2\pi i f t} dt,$$

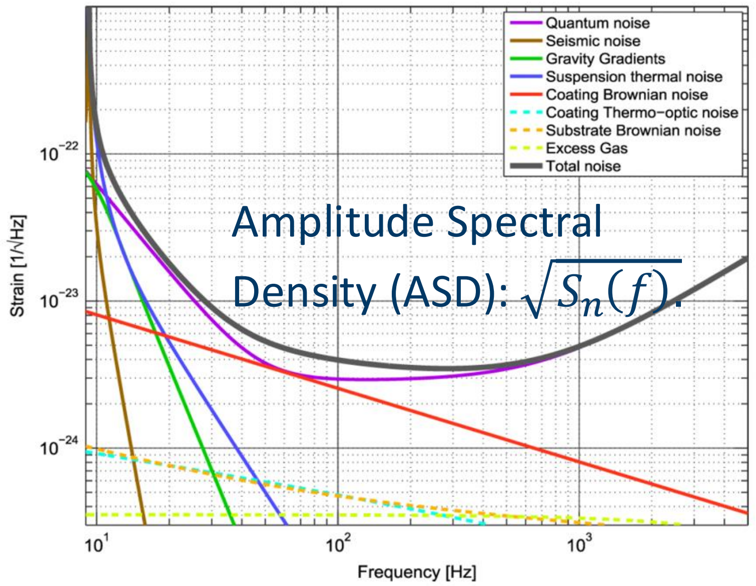
Fourier transform: time domain -> frequency domain

- Mathematically, $n(t)$ is a **stochastic process**. We don't know precisely the values of $n(t)$ so we describe it statistically: mean and correlations.
  - $n(t) = \{n(t_1), n(t_2), \dots\}$
  - Mean: zero. $< (n(t_i) > = 0,$
  - Correlation: Autocorrelation function $C(n(t_i), n(t_j)) = < n(t_i) n(t_j) >$

Noise

Simplification 1: Stationary

- Autocorrelation function $C(n(t_i), n(t_j)) = < (n(t_i), n(t_j)) >$
- Stationary: If the autocorrelation function of the noise depends only on time distance $t_i - t_j$, we say it is a **stationary** noise process
  - Noise characteristics do not change over time
  - $\tau = t_i - t_j$. The autocorrelation function becomes $C_n(\tau) =< n(t)n(t+\tau) >$
  - Wiener–Khinchin theorem: The Fourier transform of $C_n(\tau)$ is power spectral density (PSD) $S_n(f)$. $S_n(f) = \int_{-\infty}^{+\infty} C_n(\tau)\, e^{-2\pi i f \tau} d\tau$
- Real noise is not stationary! However, it is a good approximation for short duration signals



Noise

Simplification 2: Gaussian

- Autocorrelation function $C(n(t_i), n(t_j)) = < (n(t_i), n(t_j) >$
- Gaussian: The noise follows Gaussian distribution
  - Let $n_i = n(t_i), n = \{n_i\}, C_{ij} = C(n(t_i), n(t_j))$, we have
    $$P(n) = \frac{e^{-\frac{1}{2}C_{ij}^{-1}n_i n_j}}{[(2\pi)^N |\det C_{ij}|]^{1/2}},$$
  - Computing C_{ij} and its inverse is expensive.
  - Under Stationary+Gaussian assumption, noise is uncorrelated in frequency domain! Let $\tilde{n}_k = \tilde{n}(f_k), S_k = S_n(f_k)$, T=duration of data= $1/(f_{k+1} - f_k)$

$$\tilde{n}_k \sim N\left(0, \frac{TS_k}{4}\right) + iN\left(0, \frac{TS_k}{4}\right).$$

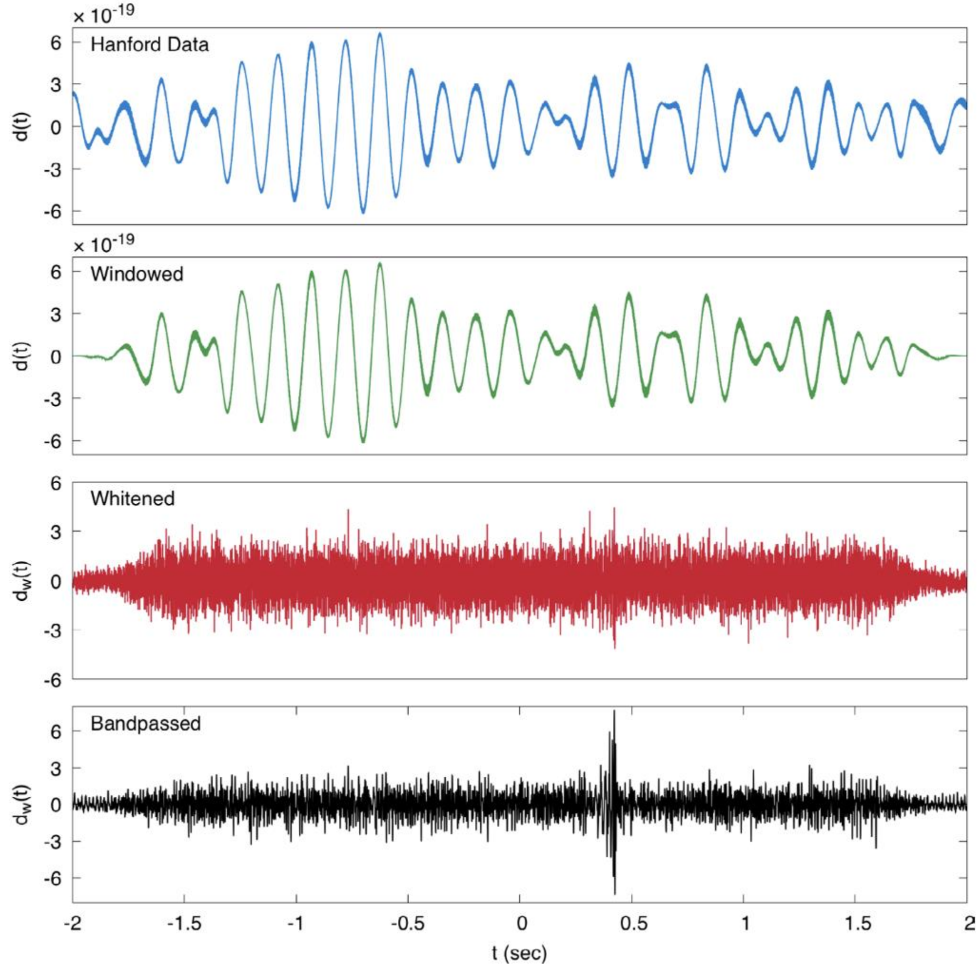
$$P(n) = \frac{1}{\prod_k \frac{\pi T S_k}{2}} e^{-\frac{1}{2}\sum_k \frac{4\tilde{n}_k \tilde{n}_k^*}{TS_k}} \propto \boxed{e^{-\frac{1}{2}(n|n)}},$$

  - Here we define **inner product** between two time/frequency series

$$(a|b) = 4\mathrm{Re}\int_0^{+\infty} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df = 4\mathrm{Re}\sum_k \frac{\tilde{a}_k^* \tilde{b}_k}{TS_k}.$$

- Gaussian assumption is far from the truth…
  - There are many transient noises (glitches)
  - Identifying and removing glitches is an important task

The data now looks like…

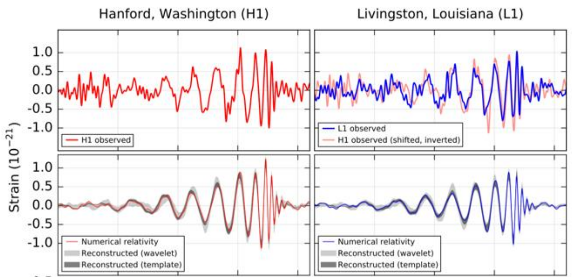


Window: Apply a window function to smooth the edge for Fourier transform

Whiten: in frequency domain, divide the data with the noise standard deviation

$$\tilde{n}_k \sim N\left(0, \frac{TS_k}{4}\right) + iN\left(0, \frac{TS_k}{4}\right)$$

Bandpass: in frequency domain, truncate the frequencies you don't want. Only target frequencies remain. In this example 50-350Hz.



Zoom in and compare with waveform template. Now you can win the Nobel pri…

…Wait, how did we get the waveform template that **matches** the data?

Detection

Matched filtering: Why

- We want to know whether signal h is present in data d or not
  - Null hypothesis H0: There is no signal present in the data. d=n
  - Alternative hypothesis H1: There is a signal h present in the data. d=h+n
- What could go wrong?
  - Type I error (missed detection): H1 is true but we mistakenly choose H0
  - Type II error (false alarm): H0 is true but we mistakenly choose H1
  - False alarm is more harmful! We want to avoid it
- Neyman-Pearson lemma: the **likelihood ratio test** is the optimal test for binary hypothesis since it minimizes the missed detection probability for a given false alarm probability

$$\mathcal{L} = \frac{p(d|\mathcal{H}_1)}{p(d|\mathcal{H}_0)} = \frac{e^{-\frac{1}{2}(d-h|d-h)}}{e^{-\frac{1}{2}(d|d)}} = e^{(d|h)}e^{-\frac{1}{2}(h|h)}.$$

  - We want to maximize the likelihood ratio
  - Since (h|h) is a constant given a template, the (d|h) is the only term to decide whether h is present, and the likelihood ratio increases monotonically with it
  - It implies that **(d|h) is relevant to the optimal detection statistic**: if it exceeds a certain threshold, we may reject H0 and assert the detection of a signal.

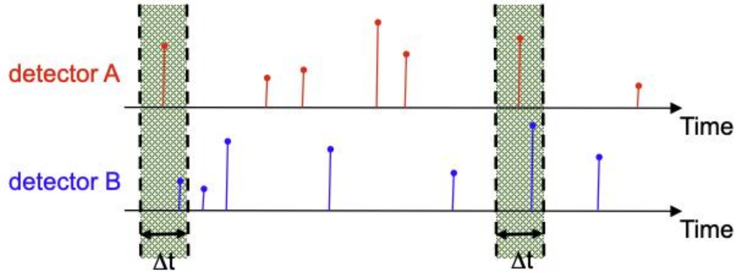
Matched filtering: How

- (d|h) is relevant to the optimal detection statistic
  - It is the inner product between the data d and template waveform h. The more similar d and h are, the larger (d|h) will be. – This is **matching** d and h.
  - From the signal processing view, (d|h) is using h as the filter function to **filter** data d. -> “Matched filtering”
- We prepare a **template bank** that contains a lot of (>1e5) templates, and **match all of them with the data**
  - Define the matched filtering signal-to-noise ratio (**SNR**): $\rho = \frac{(d|h)}{\sqrt{(h|h)}}$
  - Here (d|h) measures of the loudness of the signal, (h|h) measures the variance of the noise: $\langle (n|h)(n|h) \rangle = (h|h)$
  - The templates that give high SNR (and surpass a threshold) are the potential signal templates -> obtain initial estimates of masses, spins etc from the template
- It is proven that matched filtering (i.e., using h as the filter) is the optimal filter that maximizes the SNR defined by any filter (d|K)/√(K|K)

Other statistics

Beyond matched filtering

- Due to non-Gaussian noises, matched filtering will select much more candidates than there actually are.
- We construct other ranking statistics to further select candidates. To name a few
  - χ^2 test
    - Construct variables that follows χ^2 distribution *if* the noise is Gaussian.
    - Large χ^2 value implies non-Gaussianities in the data
  - Coincidence test
    - A true signal must arrive all detectors in a narrow time window (max dt = distance between detectors / c)



    - **False alarm rate (FAR):** shifting data from one detector so that any coincidences from possible signals can no longer occur and count coincidences again, then count how rare the new event is
  - Astrophysical origin p_{astro}
    - Given current astrophysical model for CBC and detector sensitivity, how likely the source is astrophysical?

If we are on duty today…

Trigger name	Component masses in M_sun	SNR (Usually should >8)	Chi-square $[m-\sqrt{m}, m+\sqrt{m}]$, for m+1 detectors. Assume m=2	False alarm rate (High significance: <1e-2 per year. Acceptable: <2 per year)	p_astro (usually should >0.5)	Our decision
Trigger 1	1.4+1.4	6	1.9	1e-1 per year	1.00	Not loud enough
Trigger 2	34+29	26	0.8	<1e-5 per year	0.99	A binary black hole event
Trigger 3	25+14	14	5.4	1 per month	1.00	Glitch
Trigger 4	31+1.2	9	1.3	1e-1 per year	0.49	Worth investigating

Disclaimer: 1). I made up the data. The procedure of real detection pipeline is shown in Appendix. 2). Some trigger shouldn't be called trigger because our search pipeline is smart enough to veto them. 3). In addition to data analysts, you should also consult people monitoring the detector.

Unmodeled search

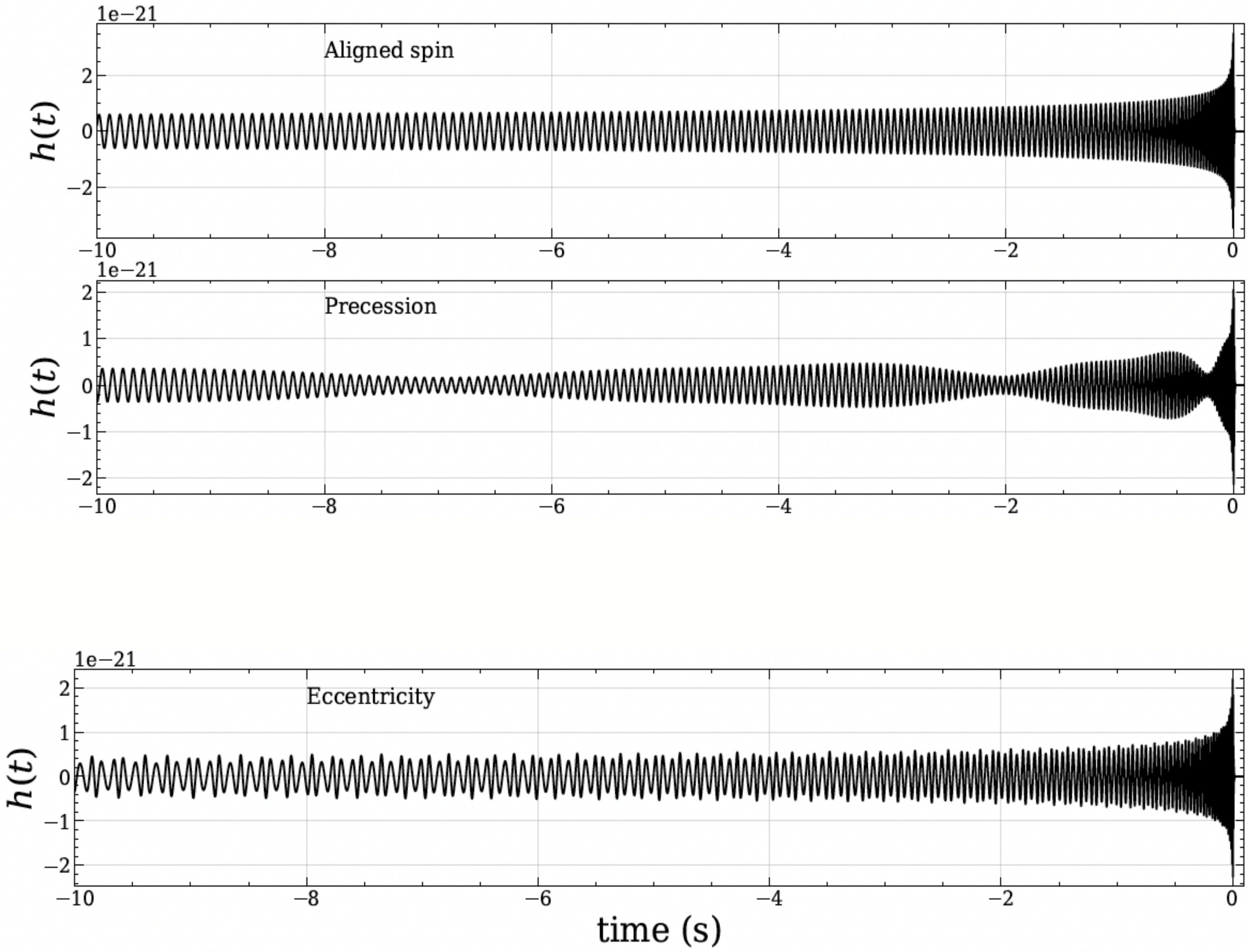
- Matched filtering only works when we have waveform templates. This is not always the case! Unmodelled signals, parameter space not covered by the template bank…
- It is possible to search for GWs only using the **coherence** between detectors: examine time and space coherence between detectors
  - Pro: unmodeled.
  - Con: Not sensitive to long signals, less sensitive than modeled search (if models are correct…), need at least 2 detectors
- In fact, the first GW detection GW150914 is detected by unmodeled search – the source (36+29 Msun) is much heavier than astrophysicists expected so they did not prepare template for that high mass…

Parameter estimation

Parametrizing compact binaries

Intrinsic parameters

- GW waveform is determined by the source properties (8+ **intrinsic parameters**)
  - Component masses (m_1, m_2, or reparametrized as chirp mass and mass ratio $\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}, q = \frac{m_2}{m_1}$)
  - Component spins ($\vec{S}_1, \vec{S}_2$, these are 6 parameters!)
  - For neutron stars, tidal deformability parameters (Λ_1, Λ_2)
  - Orbital eccentricity (e+orbital anomaly)

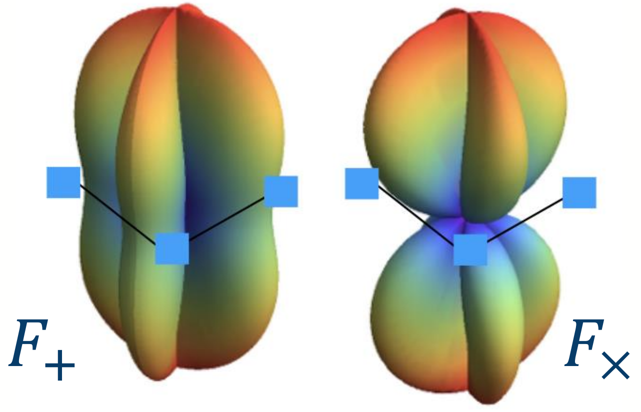


Parametrizing compact binaries

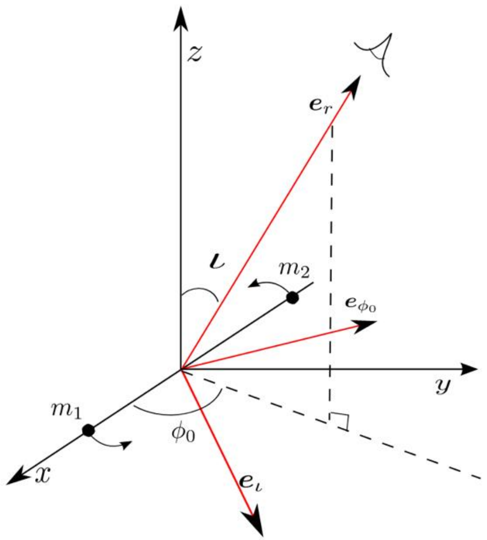
Extrinsic parameters

- GW signal is the projection of waveform polarizations to the detector:

$$h(t) = F_+ h_+(t) + F_\times h_\times(t)$$



- $F_{+,\times} = F_{+,\times}(\alpha, \delta, \psi, t_c)$ are antenna response functions
  - α, δ: right ascension, declination angles
  - ψ: polarization angle
  - t_c: coalescence time
- Other extrinsic parameters:
  - d_L: luminosity distance
  - ι: inclination angle: angle between line of sight of observer and binary orbit normal
  - ϕ_c: coalescence phase
- 7 extrinsic parameters. 15+ parameters in total!



Bayes theorem

- Parameter estimation: given observation data $\boldsymbol{d}$, we want to know the probability distribution of source parameter $\boldsymbol{\theta}$.
- Bayes theorem: we can update our knowledge after a new observation
  - Set H, the model of the source. For example, the CBC waveform model h.
  - Start from our knowledge before any observation: a **prior** distribution: e.g. source location isotropic on the sky, chirp mass uniform between 1.1 and 2 solar mass.
  - Use the observation d to update our knowledge – **likelihood**
  - **Evidence**: a normalization constant that can be ignored during parameter estimation, and reconstructed afterwards. Useful in model selection (a preferred model H should have a greater evidence)
  - The **posterior** probability represents the state of our knowledge of the model ("the truth") in light of our observed data

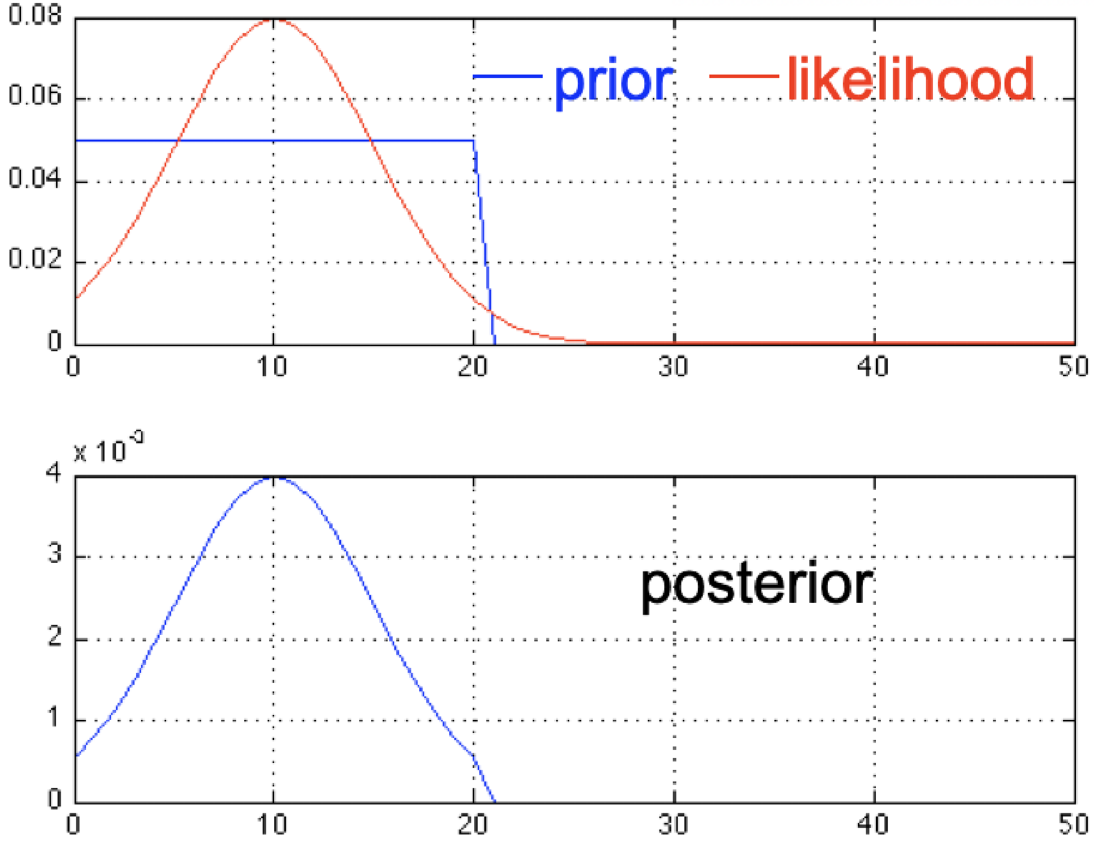
$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{H})}_{\text{Posterior}} = \frac{\overbrace{p(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{H})}^{\text{Likelihood}}\,\overbrace{p(\boldsymbol{\theta} \mid \mathcal{H})}^{\text{Prior}}}{\underbrace{p(\mathbf{d} \mid \mathcal{H})}_{\text{Evidence}}},$$

$$p(d|\theta) \propto e^{-\frac{1}{2}(n|n)} = e^{-\frac{1}{2}(d-h(\theta)|d-h(\theta))}$$

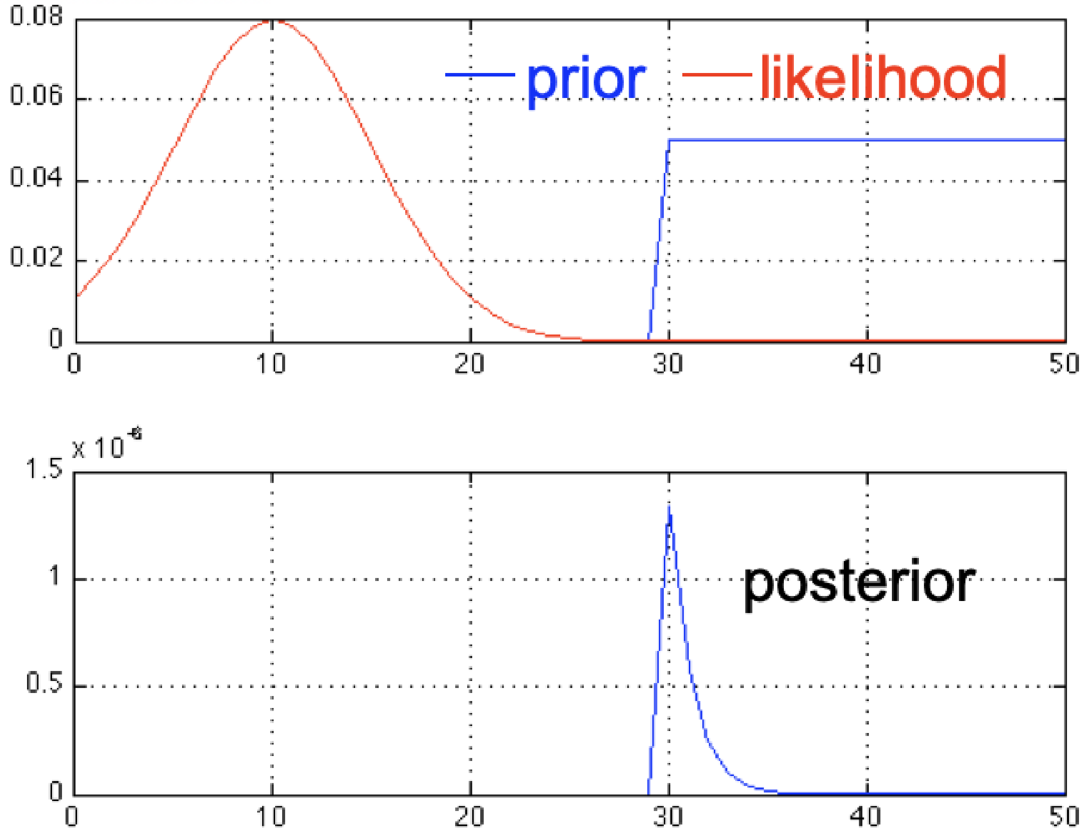
Likelihood: Subtracting the signal you get noise left

Examples

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{H})}_{\text{Posterior}} = \frac{\overbrace{p(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{H})}^{\text{Likelihood}}\,\overbrace{p(\boldsymbol{\theta} \mid \mathcal{H})}^{\text{Prior}}}{\underbrace{p(\mathbf{d} \mid \mathcal{H})}_{\text{Evidence}}},$$



Flat prior -> Posterior = Likelihood



Wrong prior -> wrong posterior

Stochastic sampling

How to obtain posterior distribution

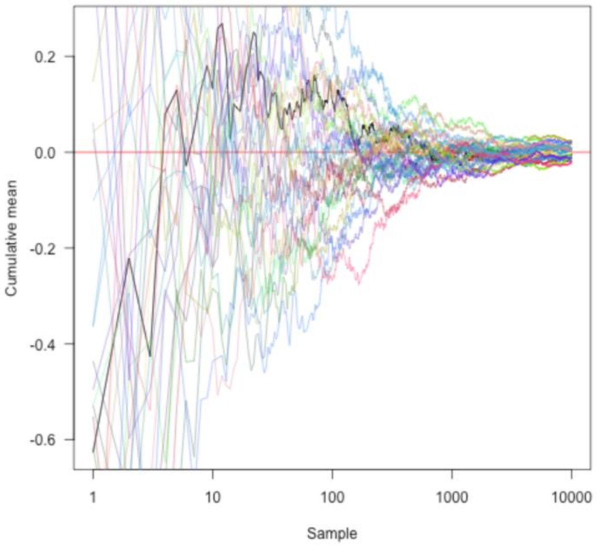
- Parameter space of CBC is at least 15 dimensions. Impossible to calculate likelihood at all places.
- **Stochastic sampling**: randomly sampling points in the parameter space (prior space), and keep high-likelihood points.
- Markov Chain Monte Carlo (**MCMC**) and **Nested sampling** are two widely used sampling algorithms.
- Very computational expensive! Takes hours – weeks.

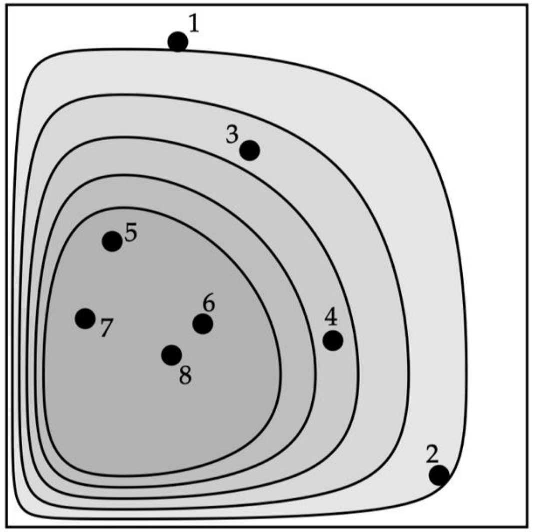
Stochastic sampling

MCMC and Nested sampling

$$\underbrace{p(\boldsymbol{\vartheta} \mid \mathbf{d}(t))}_{\text{Posterior}} = \frac{\overbrace{p(\mathbf{d}(t) \mid \boldsymbol{\vartheta})}^{\text{Likelihood}} \overbrace{p(\boldsymbol{\vartheta})}^{\text{Prior}}}{\underbrace{p(\mathbf{d}(t))}_{\text{Evidence}}},$$

- Markov Chain Monte Carlo (MCMC)
  - A set of random walkers
  - Walkers make the next step with the probability determined by target probability distribution (posterior)
    - For example, your walker is now at $\boldsymbol{\theta}$ with probability $\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{d})$
    - You randomly choose a nearby point $\boldsymbol{\theta}'$ and calculate $\mathrm{p}(\boldsymbol{\theta}'|\boldsymbol{d})$
    - Accept $\boldsymbol{\theta}'$ (i.e., let new $\boldsymbol{\theta} = \boldsymbol{\theta}'$) by the probability $\min\{1, \frac{\mathrm{p}(\boldsymbol{\theta}'|\boldsymbol{d})}{\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{d})}\}$
  - Walkers' trace will converge to posterior distribution
- Nested sampling
  - A set of live points generated from prior distributions.
  - The point with the lowest likelihood will be abandoned and the new samples with higher likelihood will be generated.
  - Repeat this until there is little new information each iteration
  - In the end, those live points will be mapped to posterior samples.

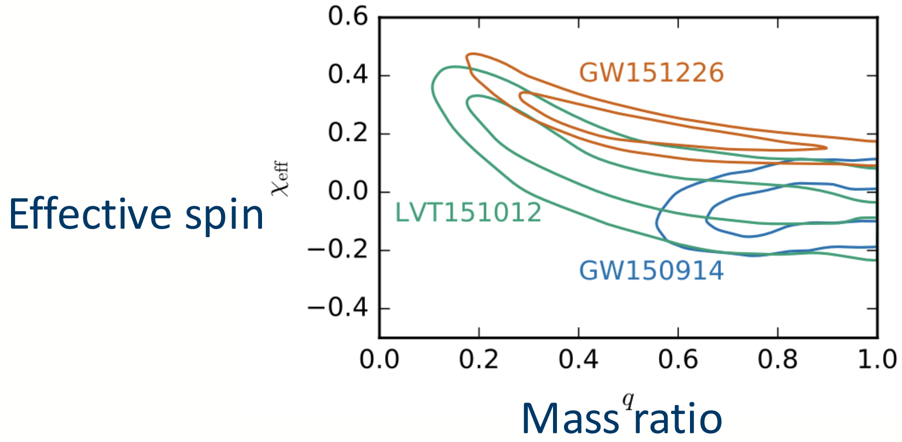




Posterior distribution

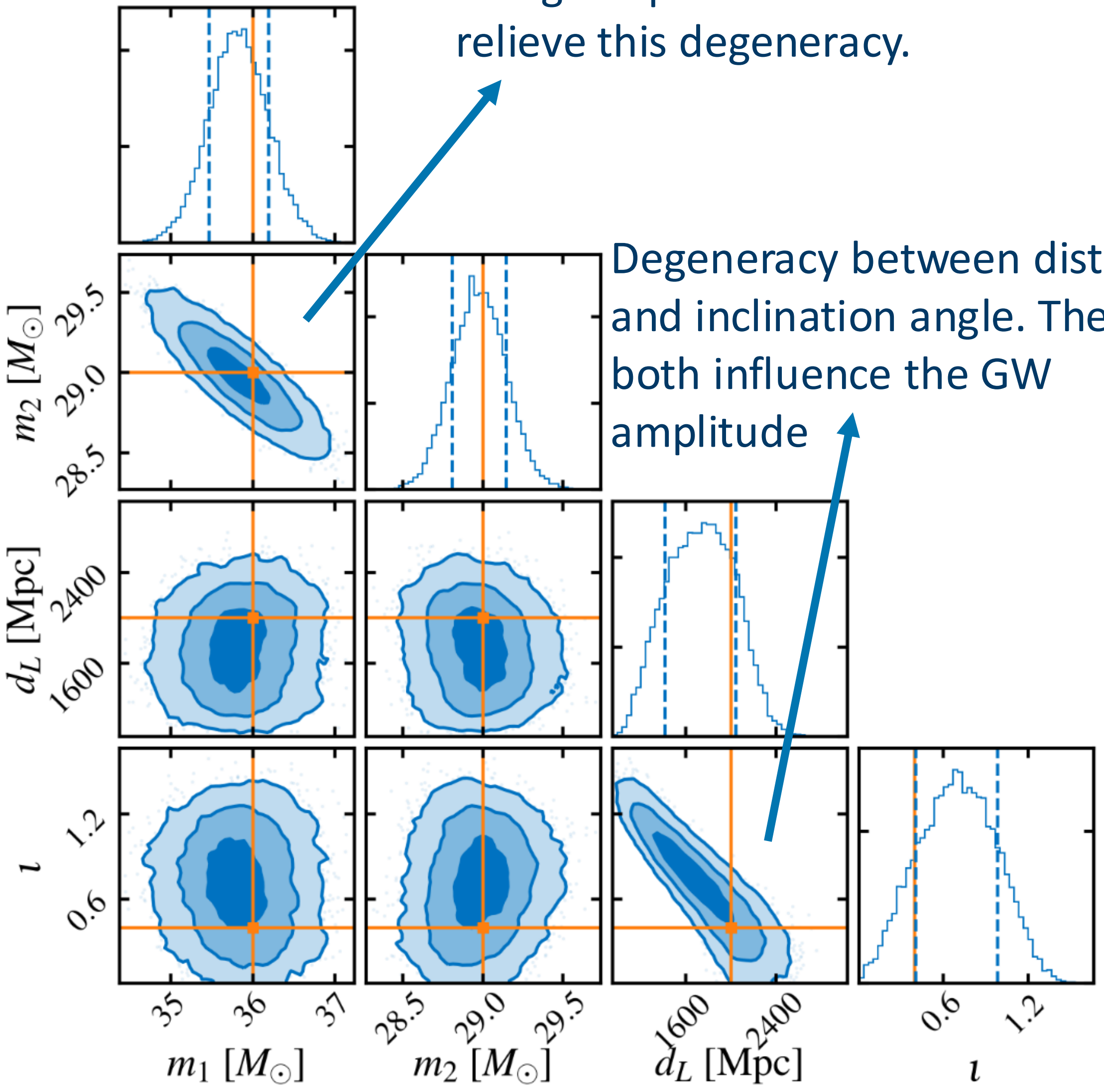
How do we interpret it?

- Visualizing the multi-dimensional distribution by corner plot
- We can read estimates of each parameters
  - How heavy are they?
  - Are they spinning?
  - Follow-up science (cosmology, population inference …)
- We can see degeneracy (correlation) between parameters



Degeneracy between m1 and m2
Using chirp mass and mass ratio can relieve this degeneracy.

Degeneracy between distance and inclination angle. They both influence the GW amplitude



Summary

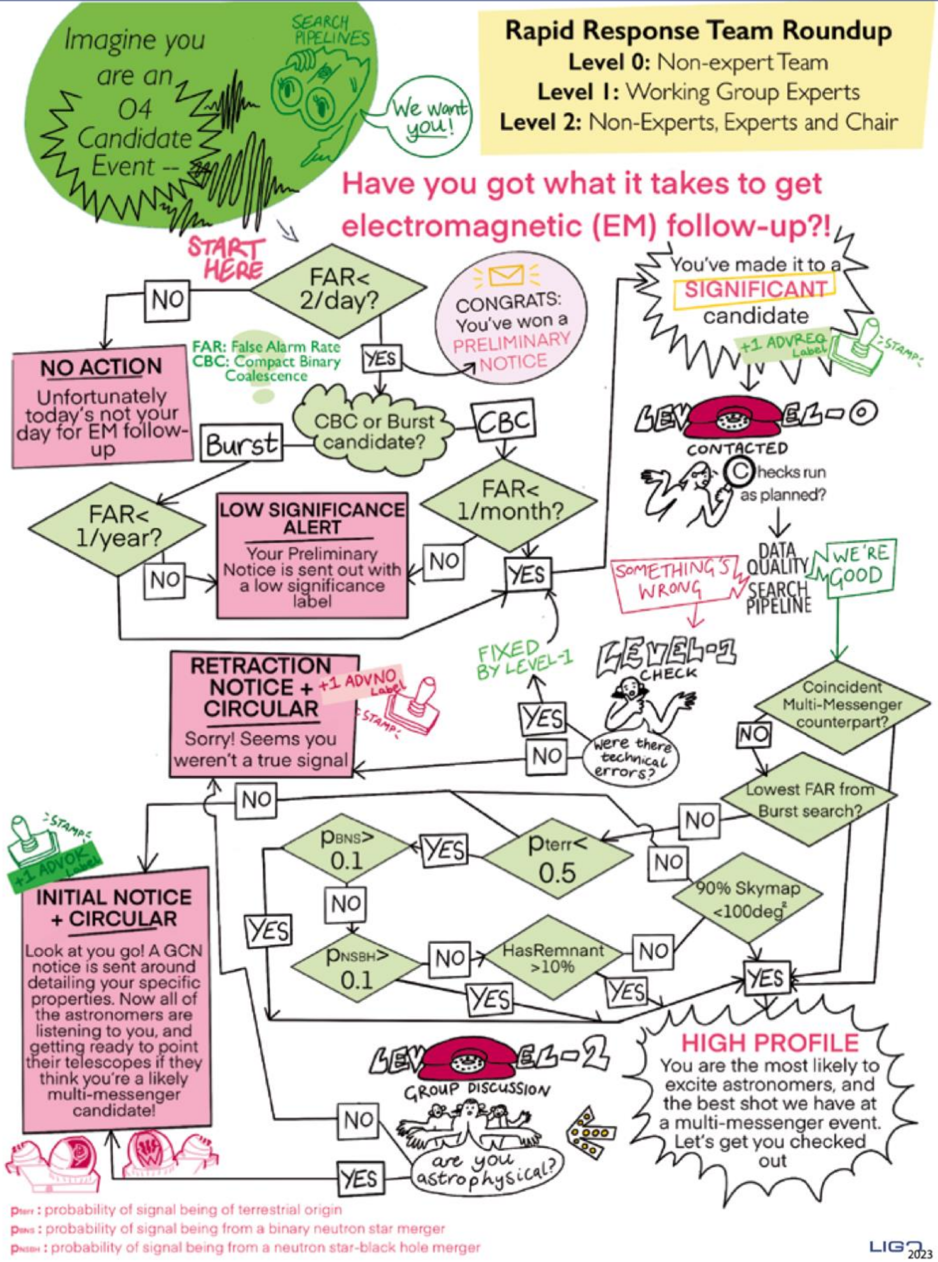
- Data analysis is the bridge connecting observation and theory
- Noise model:
  - Stationary and Gaussian assumption and when they are invalid
  - Gaussian likelihood
- Detection:
  - Matched filtering = maximum likelihood ratio, is the optimal filter
  - We need a huge template bank to perform matched filtering search
  - Other statistics is required to select real signal
  - Unmodeled search is possible
- Parameter estimation
  - Intrinsic parameters and extrinsic parameters
  - Bayesian inference and stochastic sampling
  - How to interpret parameter estimation results

What we didn't cover & further readings

1. LSC, Observation of Gravitational Waves from a Binary Black Hole Merger, https://arxiv.org/abs/1602.03837
   - First GW detection paper GW150914 (including a description of unmodeled search)
2. LVC, A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals. https://arxiv.org/abs/1908.11170
   - This is a comprehensive review, ranging from instrument to detection and parameter estimation!
   - How matched filtering search deals with extrinsic parameters (phase, time etc). Template bank actually only includes intrinsic parameters!
   - What ranking statistics are used for detection.
3. Eric Thrane, Colm Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. https://arxiv.org/abs/1809.02293
   - An excellent review of Bayesian inference in GW astronomy.
4. GW open data workshop. https://gwosc.org/s/workshop4/program.html
   - If you want to play with real data, check out GW open data workshop! Python skills required.
5. Curt Cutler and E´anna E. Flanagan, Gravitational Waves from Merging Compact Binaries: How Accurately Can One Extract the Binary's Parameters from the Inspiral Waveform? https://arxiv.org/abs/gr-qc/9402014v1
   - Fisher matrix – a powerful, widely-used tool to predict the precision of parameter estimation without running expensive stochastic sampling. We do not have time to cover this in our lecture.
6. Lee Finn, Detection, Measurement and Gravitational Radiation. https://arxiv.org/abs/gr-qc/9209010
   - Mathematical derivation of noise likelihood that we skipped in the lecture. Read if you are interested.

1,2,3,4,5 will be very useful especially if you have a research project on the relevant topic. Choose one or two to read ☺

How it works: The journey of a gravitational-wave candidate



DCC number P2300274. https://dcc.ligo.org/P2300274

 by Storm Colloms with the RRT managers: Francesco Di Renzo, Keita Kawabe, Hisaaki Shinkai, Peter Shawhan