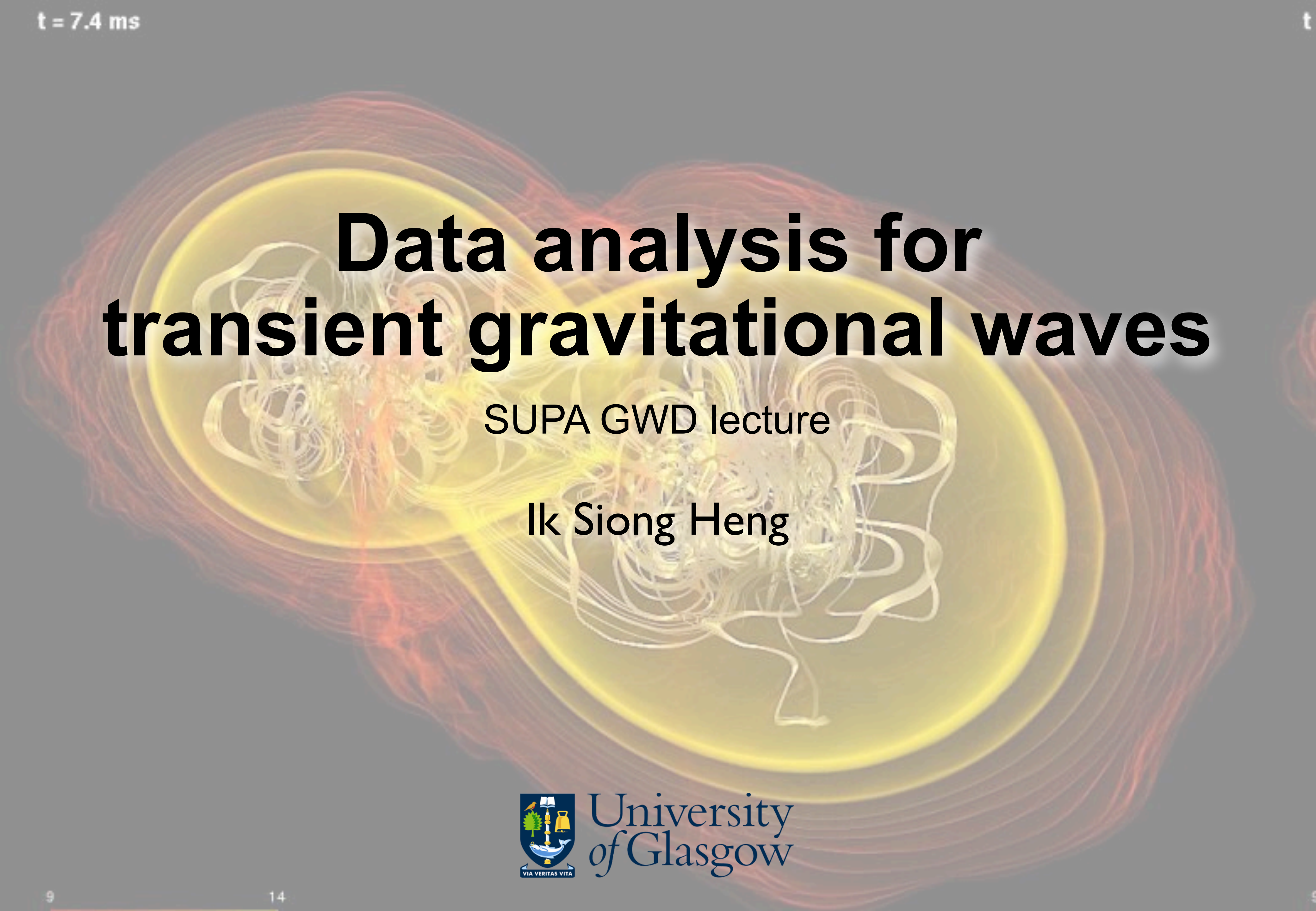


Data analysis for transient gravitational waves

SUPA GWD lecture

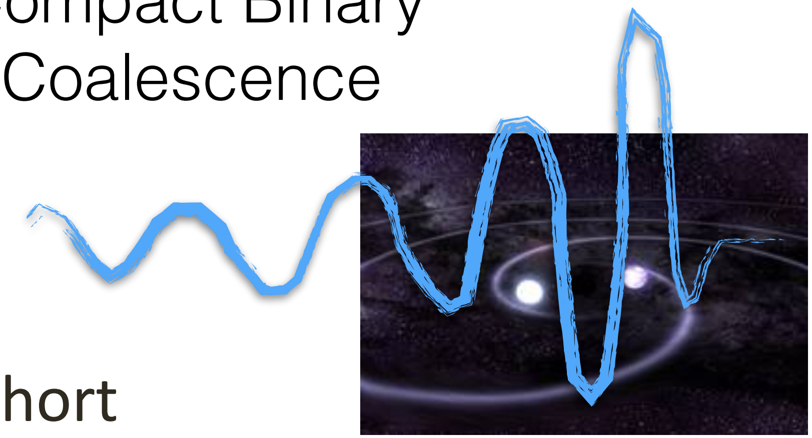
Ik Siong Heng

University of Glasgow

Gravitational wave sources types

modelled	unmodelled
Compact Binary Coalescence	Burst
Continuous	Stochastic

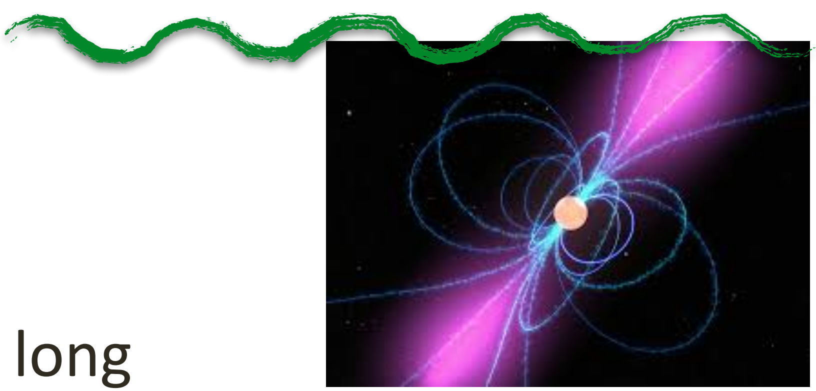
Compact Binary Coalescence



short

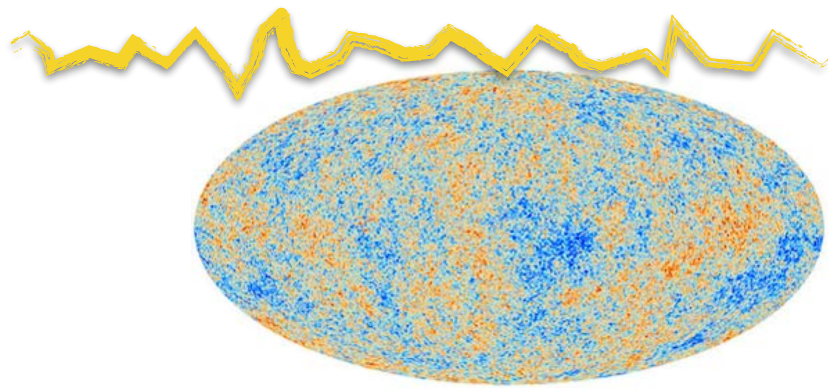
Burst

Continuous



long

Stochastic



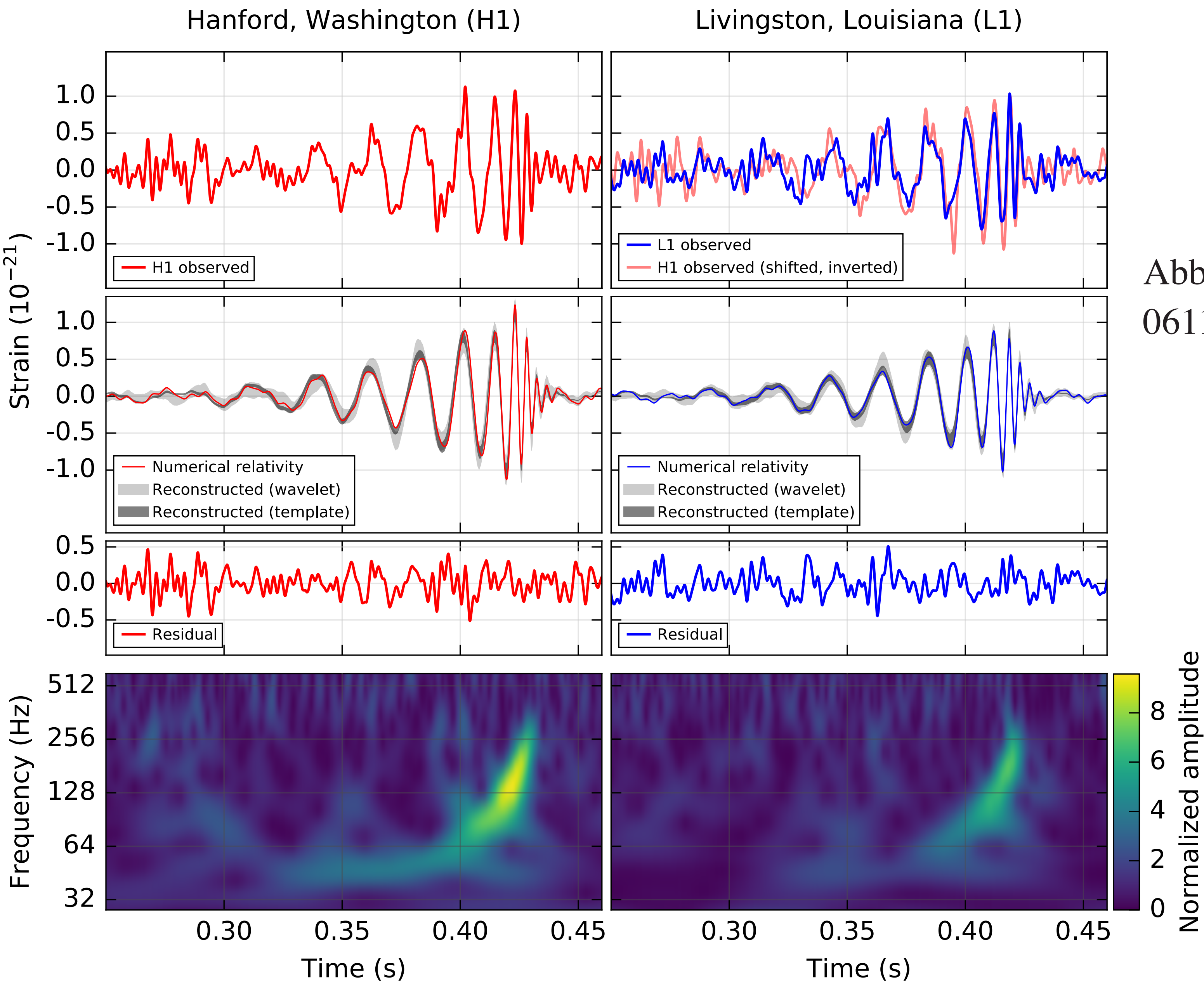
Search methods

- **Two kinds of burst and CBC searches**
  - -all-sky search: maximise detection efficiency across the whole sky
  - -triggered search: search trigged by non-GW observation (eg. GRBs)

- **Calculate ranking statistic**
  - -ranking statistic determined by two main approaches
    - ➡coherent searches: combine data from multiple detectors
    - ➡coincidence searches: identify triggers in each detector and perform coincidence

- **Apply a threshold after processing to pick out the largest SNR event**

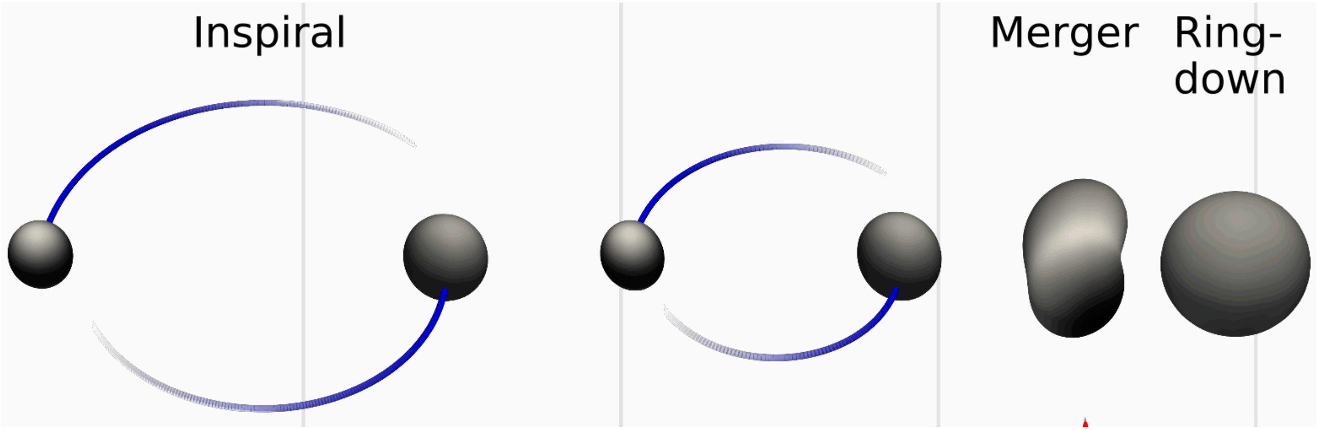
First detection



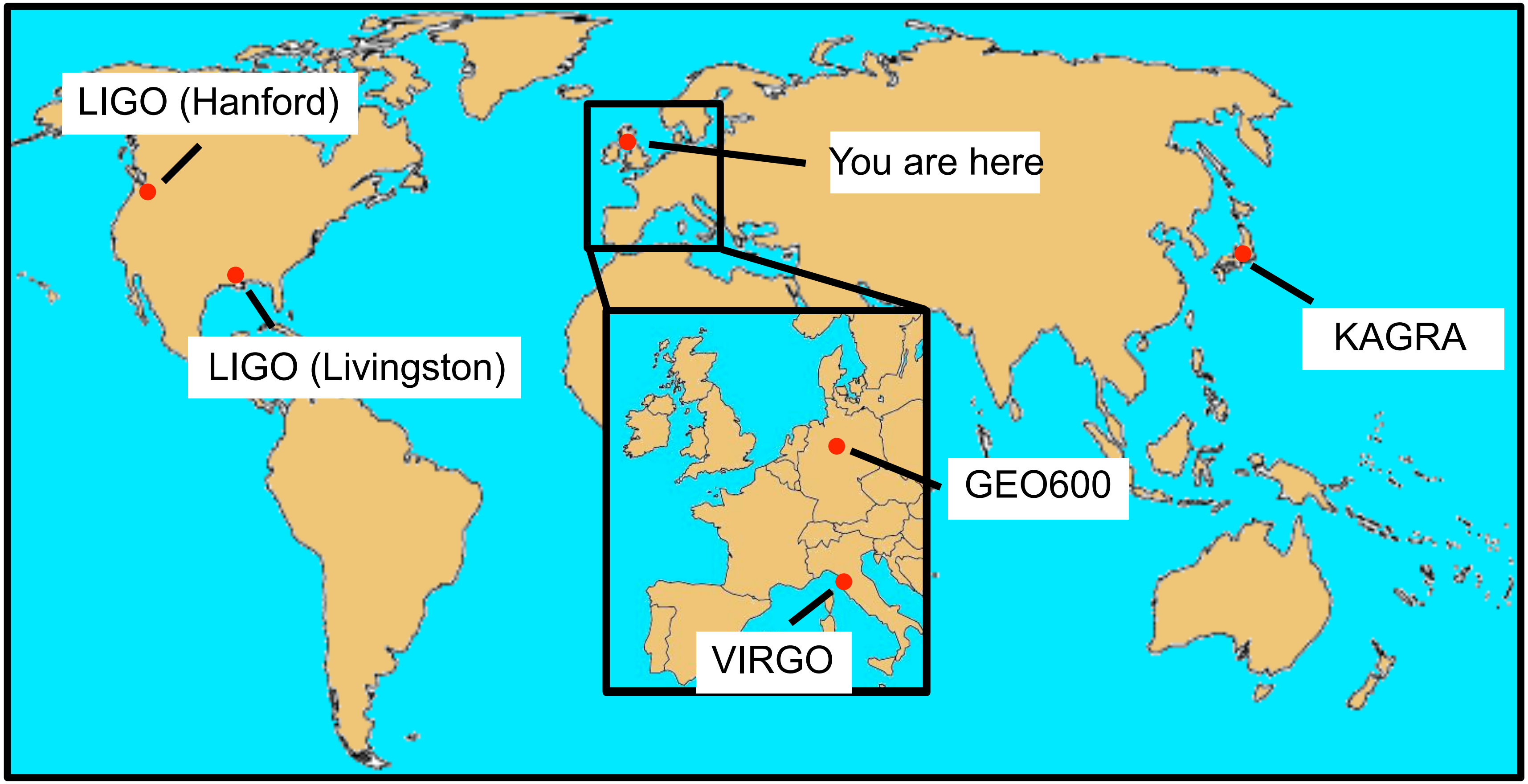
Abbot et al., PRL 116, 061102 (2016)

Detection by generic transient searches

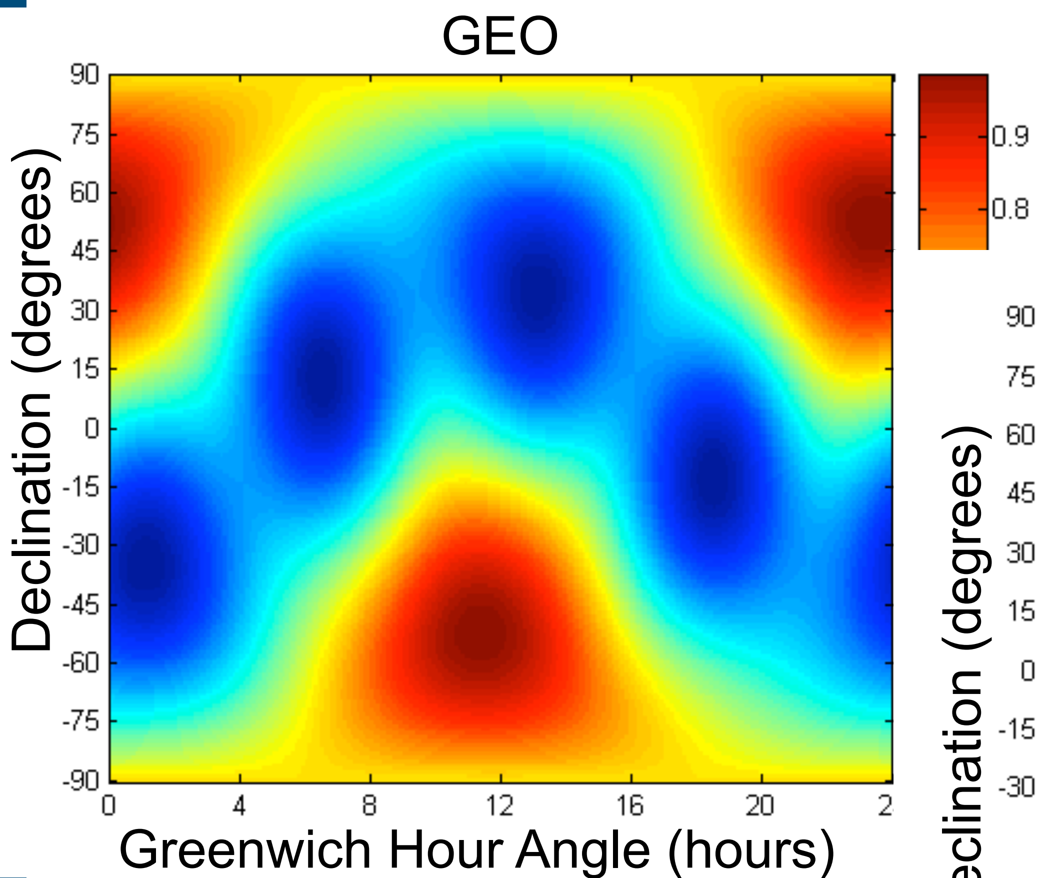
- Excerpt from the detection paper:
  "On September 14, 2015 at 09:50:45 UTC, the LIGO Hanford, WA, and Livingston, LA, observatories detected the coincident signal GW150914 shown in Fig. 1. The initial detection was made by low-latency searches for generic gravitational-wave transients [41] and was reported within three minutes of data acquisition [43]."



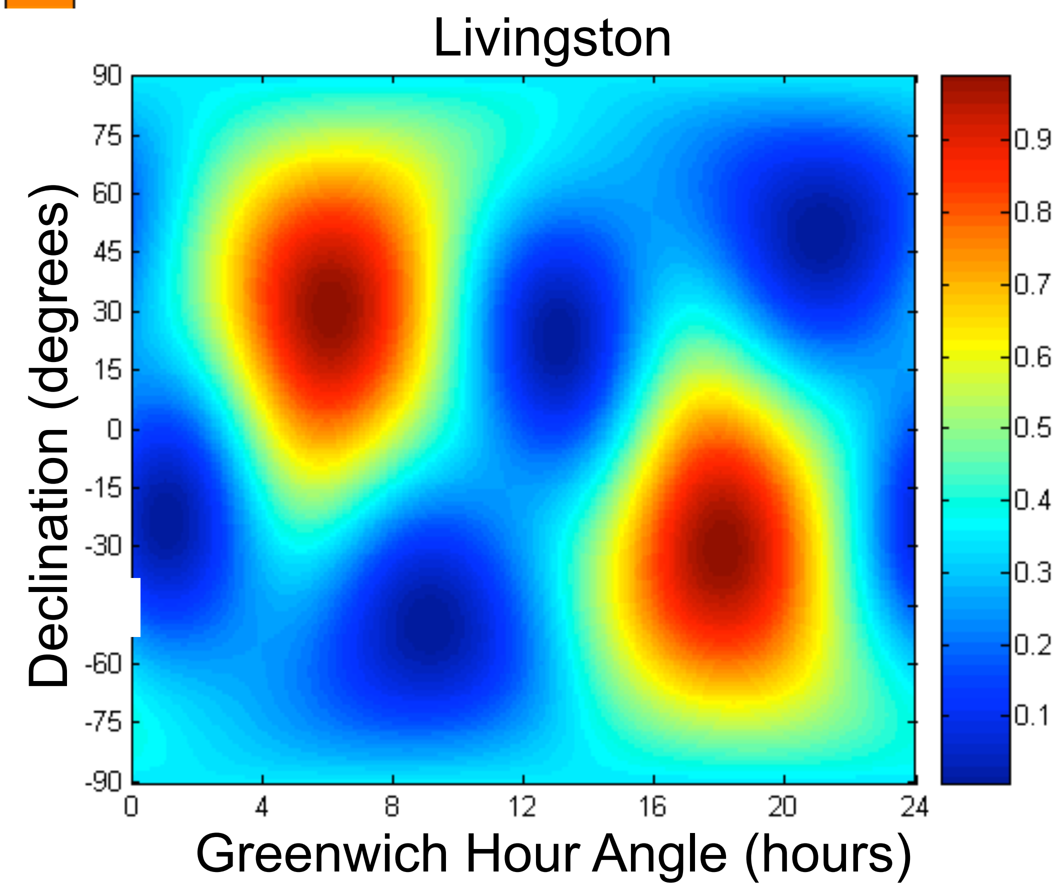
Interferometric detectors worldwide



Directional sensitivity



$F_+^2 + F_x^2$



- Lets formulate our data such that

$$\tilde{\boldsymbol{d}}_{\mathrm{w}} = \boldsymbol{F}_{\mathrm{w}}\tilde{\boldsymbol{h}} + \tilde{\boldsymbol{n}}_{\mathrm{w}},$$

or

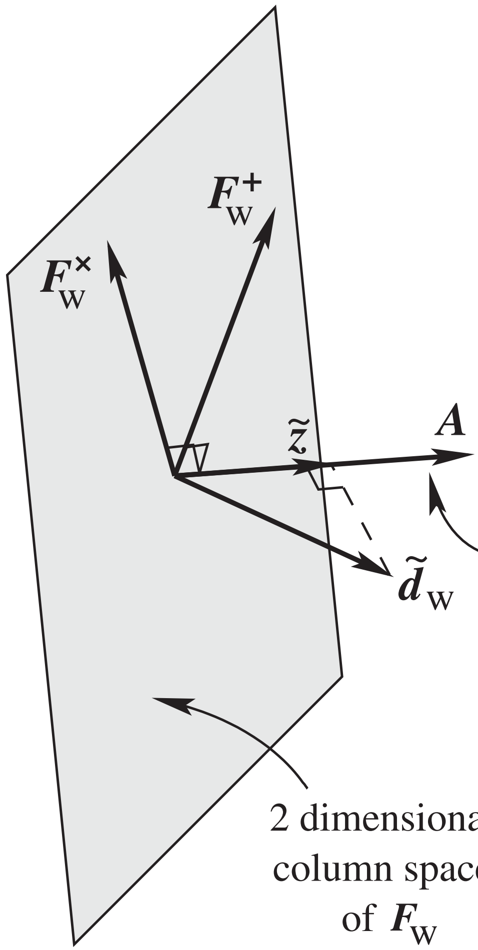
data, antenna pattern, signal, noise

$$\begin{bmatrix} \tilde{d}_{w1} \\ \tilde{d}_{w2} \\ \vdots \\ \tilde{d}_{wD} \end{bmatrix} = \begin{bmatrix} F^{+}_{w1} & F^{\times}_{w1} \\ F^{+}_{w2} & F^{\times}_{w2} \\ \vdots & \vdots \\ F^{+}_{wD} & F^{\times}_{wD} \end{bmatrix} \begin{bmatrix} \tilde{h}_{+} \\ \tilde{h}_{\times} \end{bmatrix} + \begin{bmatrix} \tilde{n}_{w1} \\ \tilde{n}_{w2} \\ \vdots \\ \tilde{n}_{wD} \end{bmatrix},$$

sky location, frequency

$$\boldsymbol{F}_{\mathrm{w}}(\hat{\Omega}_s, f) \equiv [\,\boldsymbol{F}^{+}_{\mathrm{w}} \quad \boldsymbol{F}^{\times}_{\mathrm{w}}\,] = \begin{bmatrix} F^{+}_{\mathrm{w}1} & F^{\times}_{\mathrm{w}1} \\ F^{+}_{\mathrm{w}2} & F^{\times}_{\mathrm{w}2} \\ \vdots & \vdots \\ F^{+}_{\mathrm{w}D} & F^{\times}_{\mathrm{w}D} \end{bmatrix}$$

Chatterji et al, PRD 74, 082005 (2006)



- **Consider a matrix A whose rows are components of an orthonormal basis**
- **If we construct such that**

$$\boldsymbol{A}\,\boldsymbol{F}_{\mathrm{w}} = 0.$$

- **then, we can use it to construct a null stream**

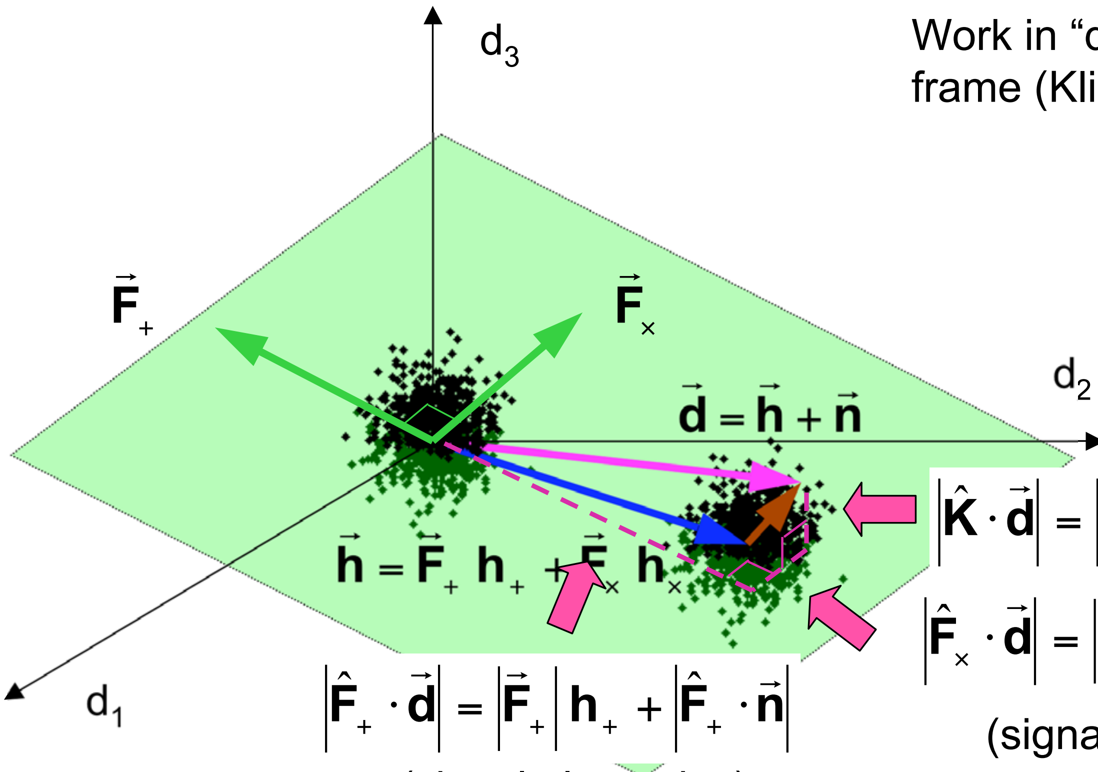
$$\tilde{\boldsymbol{z}} \equiv \boldsymbol{A}\tilde{\boldsymbol{d}}_{\mathrm{w}} = \boldsymbol{A}\boldsymbol{F}_{\mathrm{w}}\tilde{\boldsymbol{h}} + \boldsymbol{A}\tilde{\boldsymbol{n}}_{\mathrm{w}} = \boldsymbol{A}\tilde{\boldsymbol{n}}_{\mathrm{w}}.$$

- **For 3 detectors, A is constructed by**

$$\boldsymbol{A} = \frac{\boldsymbol{F}_{\mathrm{w}}^{+} \times \boldsymbol{F}_{\mathrm{w}}^{\times}}{|\boldsymbol{F}_{\mathrm{w}}^{+} \times \boldsymbol{F}_{\mathrm{w}}^{\times}|}$$

Chatterji et al, PRD 74, 082005 (2006)

Geometric View: One TF Pixel



Work in "dominant polarization" frame (Klimenko et al, PRD '05):

$$\vec{F}_+ \cdot \vec{F}_\times = 0$$

$$\left|\hat{K} \cdot \vec{d}\right| = \left|\hat{K} \cdot \vec{n}\right| \quad \text{(noise only)}$$

$$\left|\hat{F}_\times \cdot \vec{d}\right| = \left|\vec{F}_\times\right| h_\times + \left|\hat{F}_\times \cdot \vec{n}\right|$$

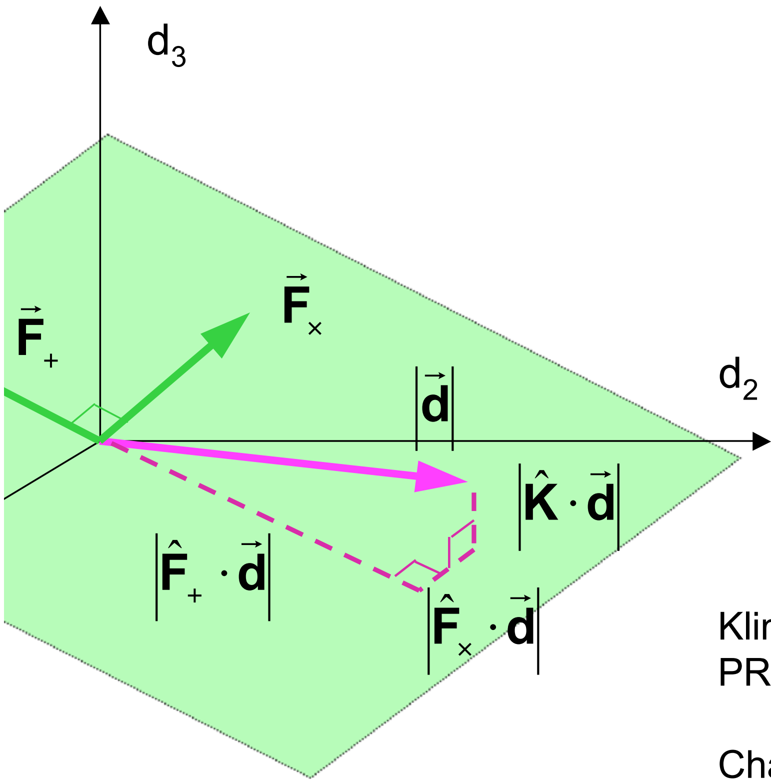
(signal plus noise)

$$\left|\hat{F}_+ \cdot \vec{d}\right| = \left|\vec{F}_+\right| h_+ + \left|\hat{F}_+ \cdot \vec{n}\right|$$

(signal plus noise)

$$\hat{K} \equiv \hat{F}_+ \times \hat{F}_\times$$

Likelihood or Energy Measures



Likelihoods used in this analysis:

Hard Constraint $E_{HC} = \left|\hat{F}_+ \cdot \vec{d}\right|^2$

Null Energy $E_{NULL} = \left|\hat{K} \cdot \vec{d}\right|^2$

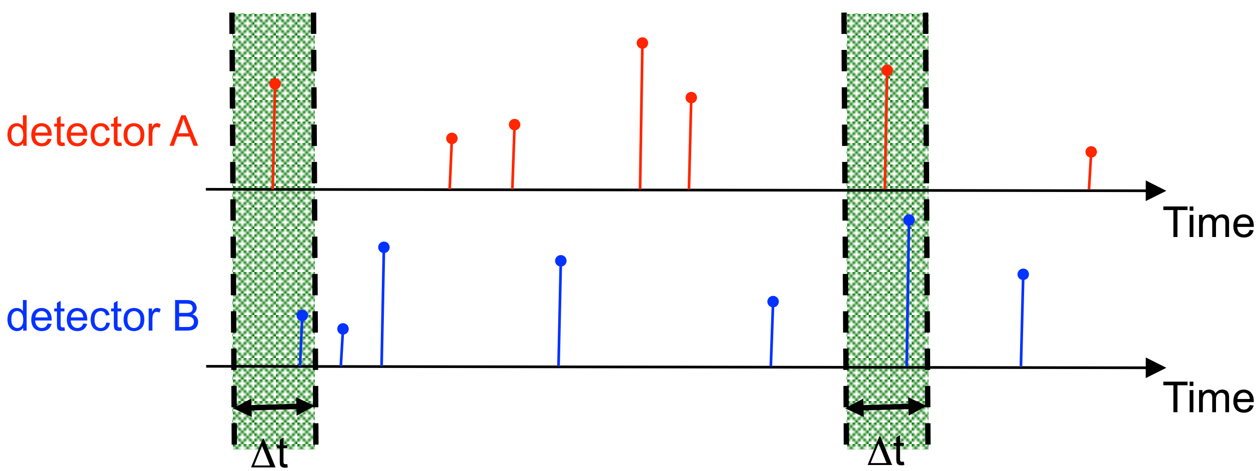
Incoherent Energy $E_{INC} = \sum_\alpha \left|K_\alpha \cdot d_\alpha\right|^2$

Klimenko, Mohanty, Rakhmanov,& Mitselmakher,
PRD **72** 122002 (2005); J. Phys. Conf. Ser. **32** 12 (2006)

Chatterji, Lazzarini, Stein, Sutton, Searle, & Tinto,
PRD **74** 082005 (2006)

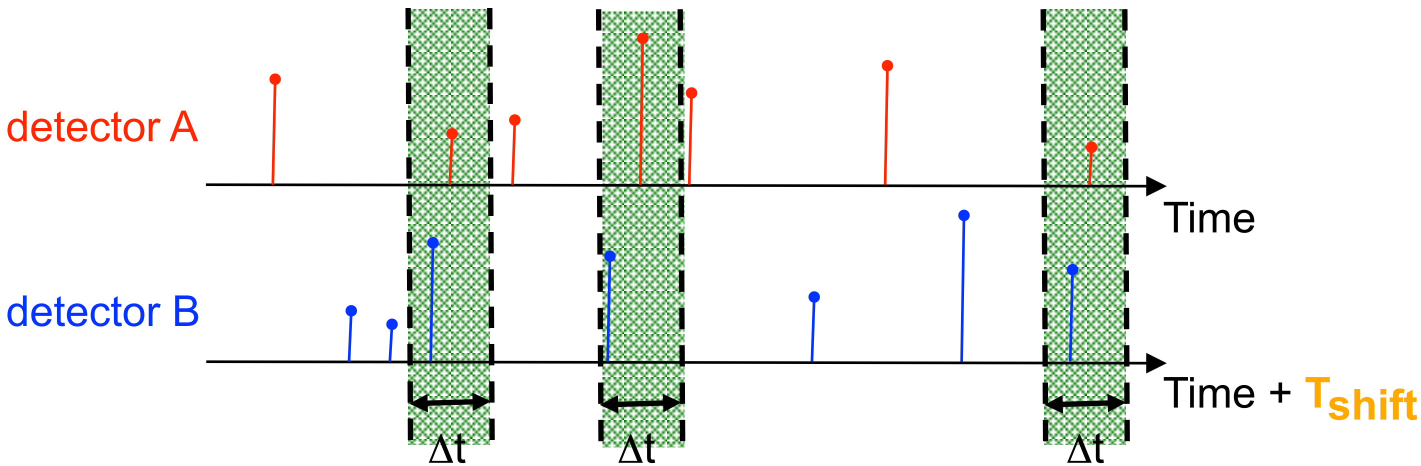
Coincidence analysis

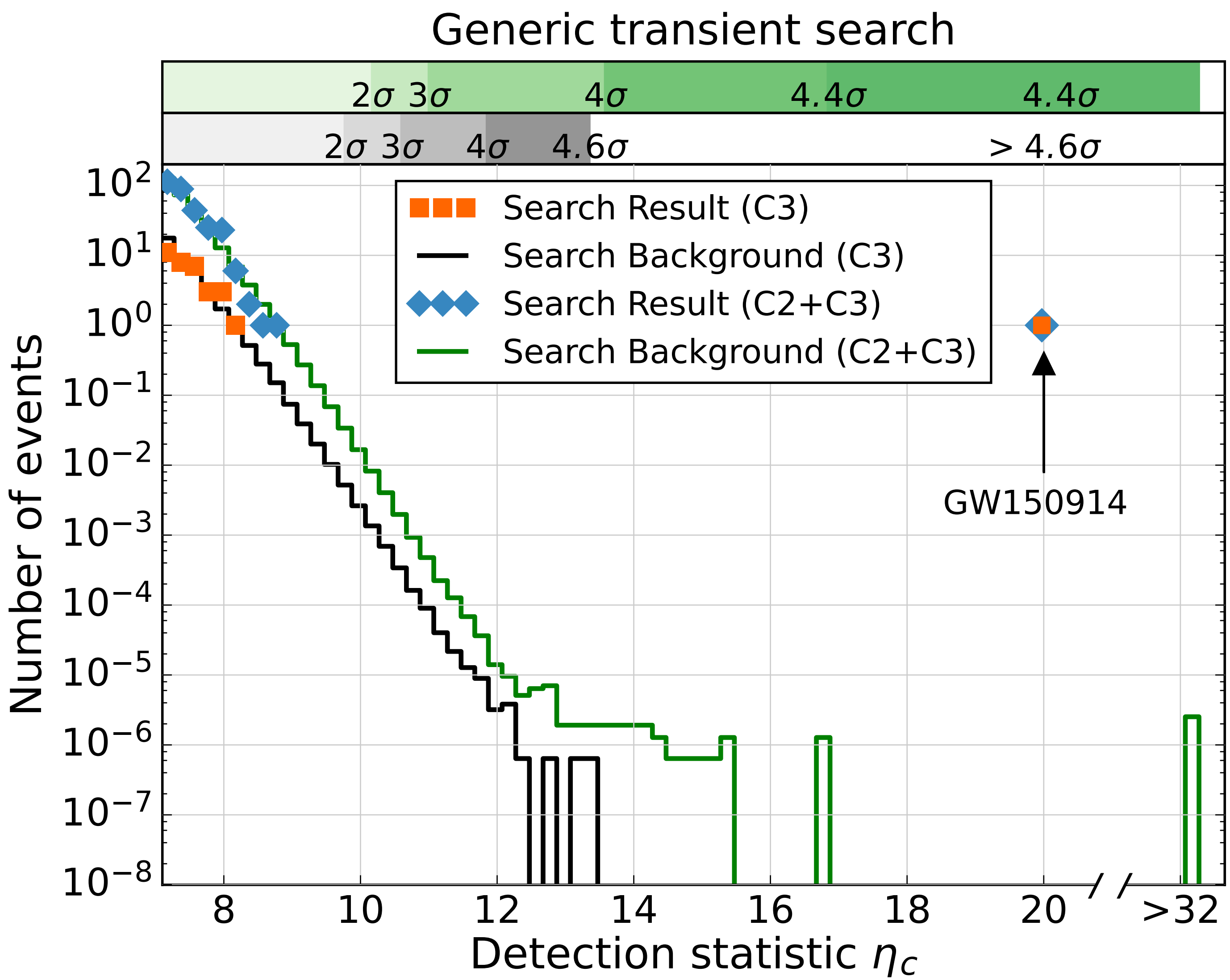
- **Burst gravitational wave will** excite 2 widely-spaced detectors almost simultaneously (within light travel time between the detectors)
- **To identify possible signal,** count number of events that are coincident within a particular time window, Δt



Coincidence analysis

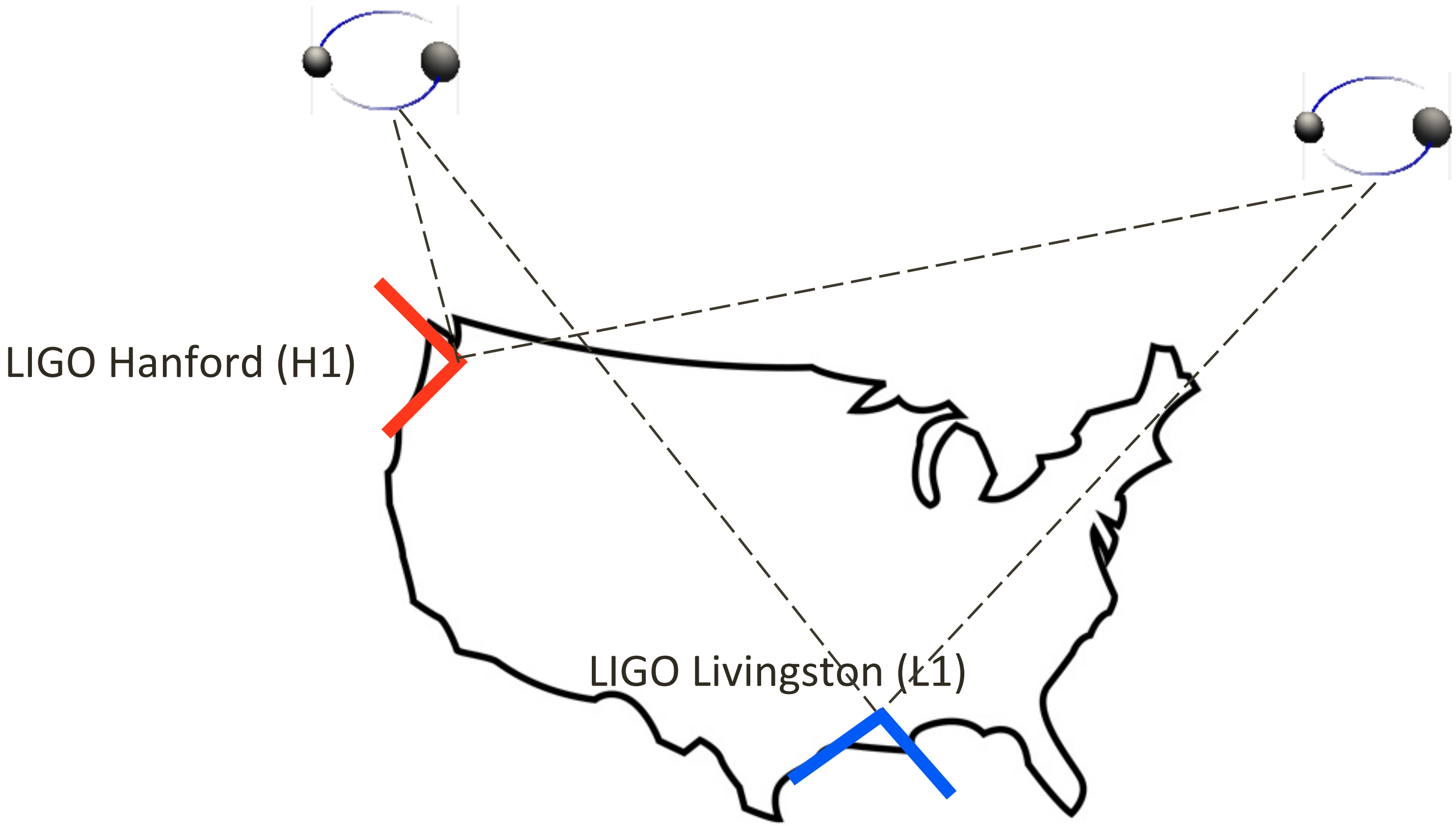
- **Need to** estimate the rate of "accidental" coincidences (false alarms)
- **Add a time shift** (T_{shift}) to the data so that any coincidences from possible signals can no longer occur and count coincidences again



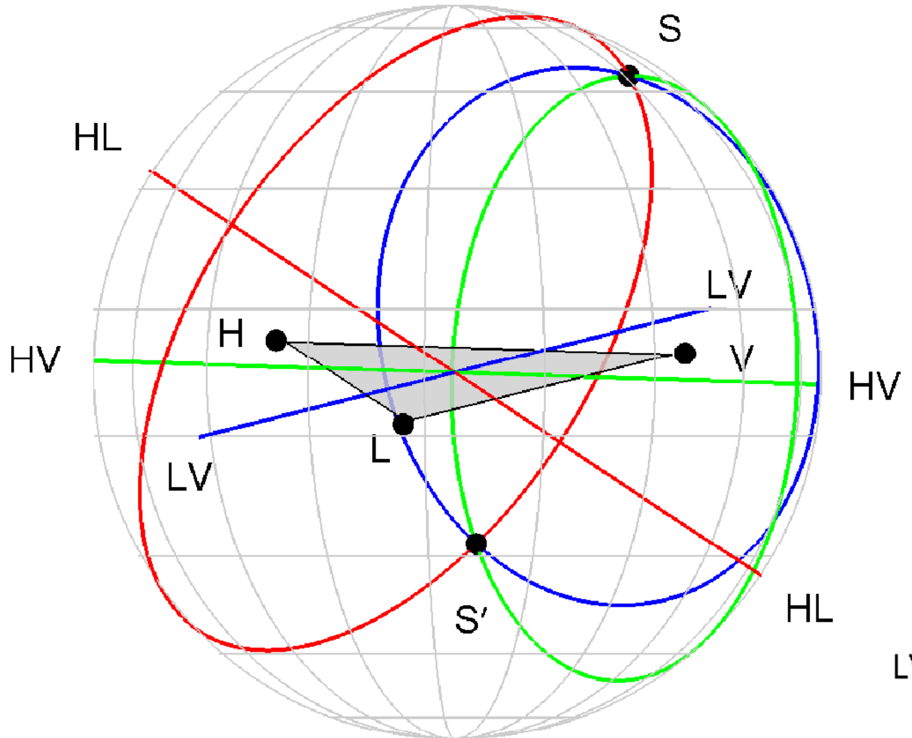

Generic transient search
2σ 3σ
4σ
4.4σ
4.4σ
2σ 3σ
4σ
4.6σ
> 4.6σ
Search Result (C3)
Search Background (C3)
Search Result (C2+C3)
Search Background (C2+C3)
GW150914
Number of events
Detection statistic η_c
>32

Rapid sky localisation

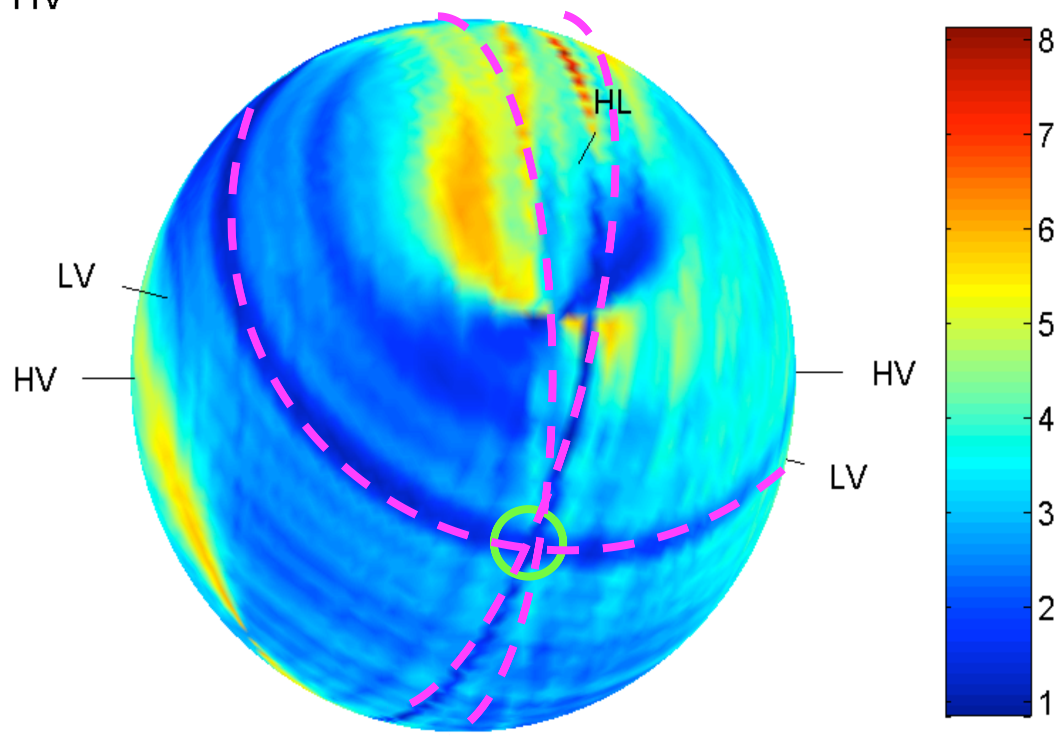
- The sky location of the gravitational wave source can be estimated through the event arrival time at each detector
  - ➡gravitational waves travel at the speed of light
- Rapid sky localisation allows search for counterpart transient signals by optical, X-ray, GRB, ... observatories



Source sky location

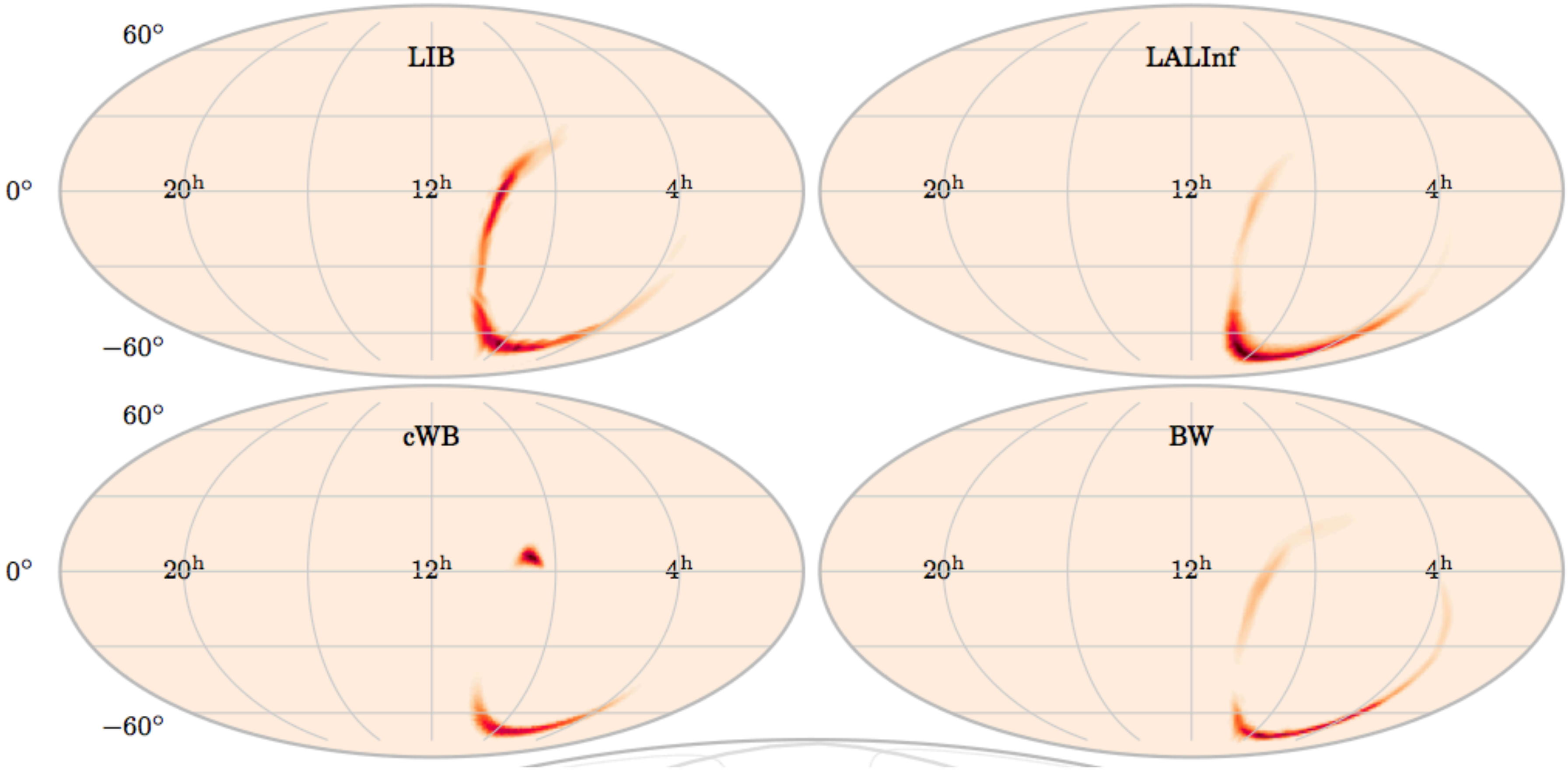


H - LIGO Hanford
L - LIGO Livingston
V - Virgo



Use time delay between detectors to estimate sky location of GW signal source

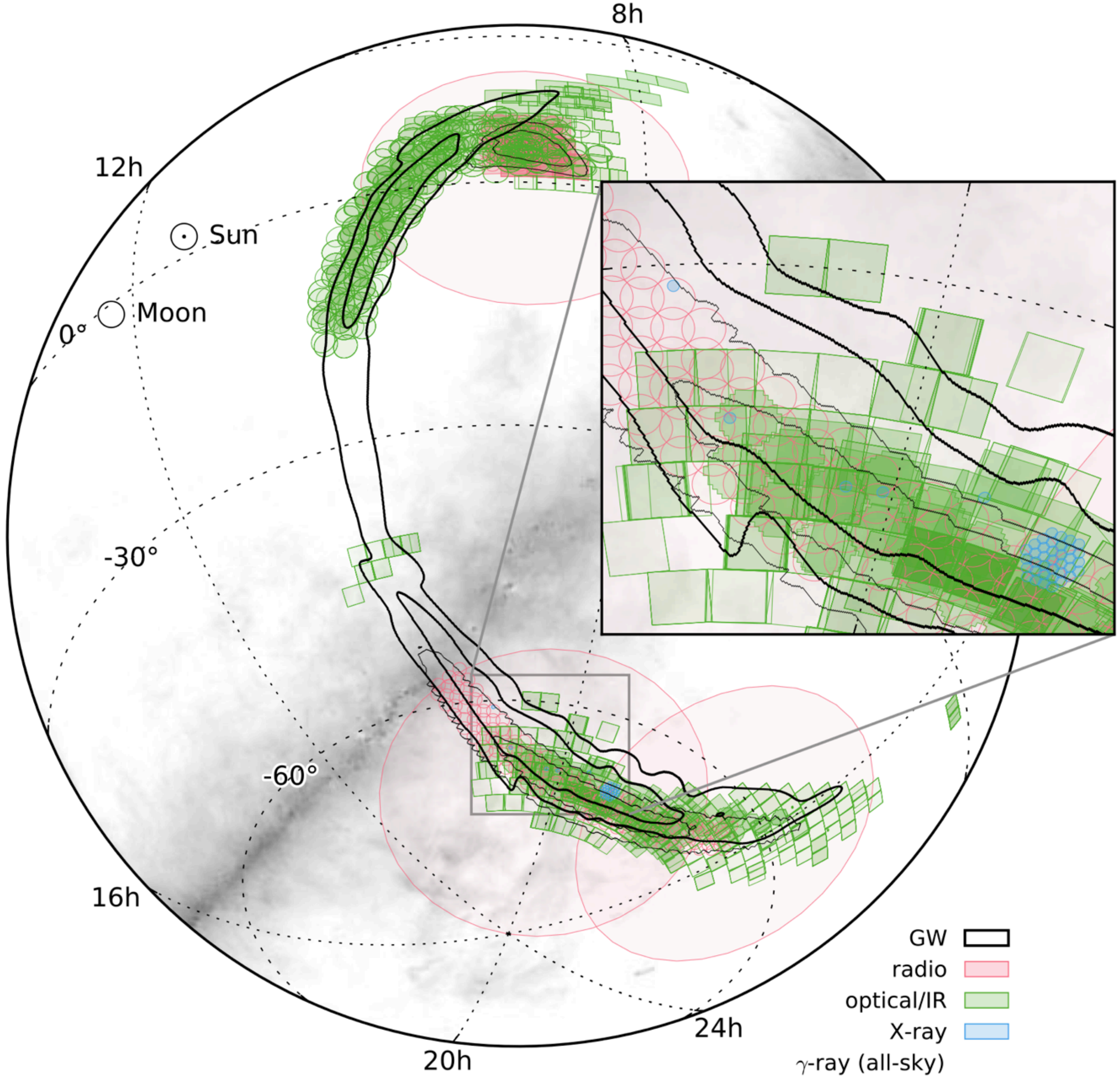
Rapid sky localisation



- cWB, LIB, BW: burst analyses
- LALInf: sky location estimate assuming full waveform model

http://arxiv.org/abs/1602.03843

Follow-up of GW150914



Compact binary coalesce signal

- Compact binary systems are binaries systems which consist of neutron stars and/or black holes
- As the compact objects orbit, angular momentum is radiated away as gravitational radiation, leading to an inspiralling orbit
- The gravitational wave signal for the two polarisations take the form

$$h_+(t) = A_{\mathrm{GW}}(t)(1+\cos^2\iota)\cos\phi_{\mathrm{GW}}(t)$$

$$h_\times(t) = -2A_{\mathrm{GW}}(t)\cos\iota\sin\phi_{\mathrm{GW}}(t)$$

- The observed signal is a function of the antenna patterns

$$h_{\mathrm{det}}(t) = F_+h_+(t) + F_\times h_\times(t)$$

Matched filtering

- We follow the treatment in the review paper

http://relativity.livingreviews.org/Articles/lrr-2009-2/fulltext.html

- The data is a combination of the signal and nose

$$x(t) = h(t - t_a) + n(t)$$

- Since the phase evolution of the signal is well modelled, the strategy is to correlate the data with templates of the expected signal, ***q*(*f*)**,

$$c(\tau) = \int_{-\infty}^{\infty} \tilde{x}(f)\tilde{q}^*(f)e^{-2\pi i f \tau} df$$

where we are working with the fourier transforms:

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t)e^{2\pi i f t} df$$

Matched filtering

- If we cross correlate a signal with its corresponding template, we obtain a value that reflects the strength of the signal

$$\Sigma \equiv \bar{c}(\tau) = \int_{-\infty}^{\infty} \tilde{h}(f)\tilde{q}^*(f)e^{-2\pi i f\tau}df$$

- If we subtract the signal from the data and correlate, we obtain a measure of the noise power

$$N^2 = \overline{(c-\bar{c})^2} = \int_{-\infty}^{\infty} S_h(f)|\tilde{q}^*(f)|^2 df$$

- $S_h(f)$ is noise power spectral density and is a measure of the variance of the noise at frequency
- Thus, the square of the signal-to-noise ratio is $\rho^2 = \frac{\Sigma^2}{N^2}$

Optimal signal-to-noise ratio

- **We introduce noise-weighted inner products**

$$\langle a, b\rangle \equiv 2\int_0^\infty \frac{df}{S_h(f)}[\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)]$$

- **The square of the signal-to-noise ratio becomes**

$$\rho^2 = \frac{\langle h e^{2\pi i f(\tau - t_A)}, S_h q\rangle}{\sqrt{\langle S_h q, S_h q\rangle}}$$

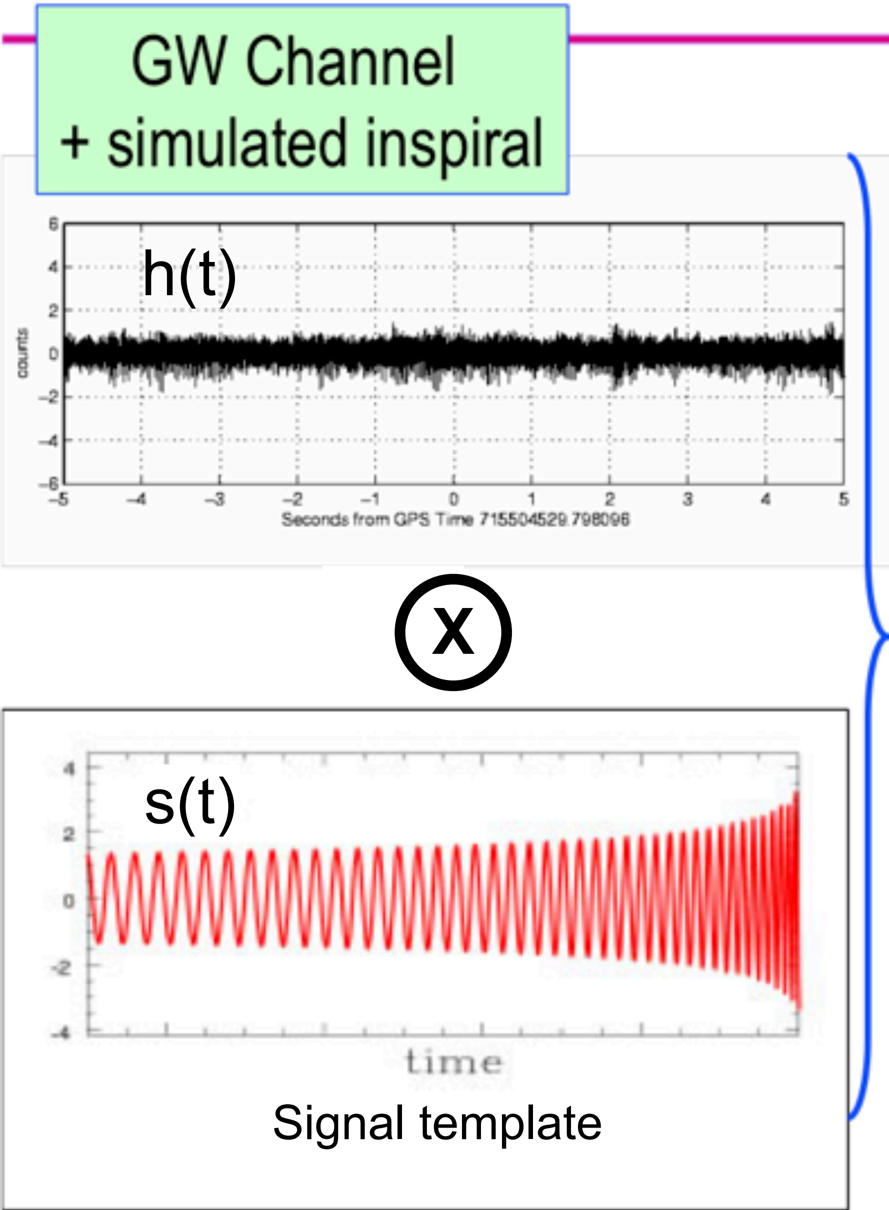
- **The ρ^2 is maximised when**

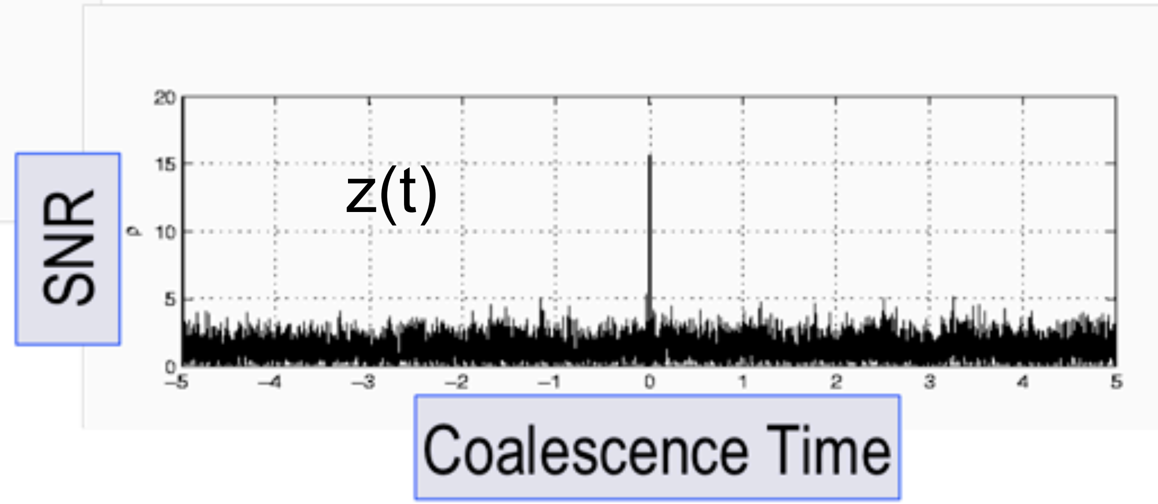
$$\tilde{q}(f) \propto \frac{\tilde{h}(f) e^{2\pi i f(\tau - t_a)}}{S_h(f)}$$

- **Substituting this, we find**

$$\rho^2_{\text{opt}} = 4\int_0^\infty \frac{|\tilde{h}(f)|^2}{S_h(f)} df$$

Calculating Inspiral SNR





$$z(t) = 4 \int_0^\infty \frac{\tilde{h}(f)^* \tilde{s}(f)}{S_n(f)} e^{2\pi i f t} \, df,$$

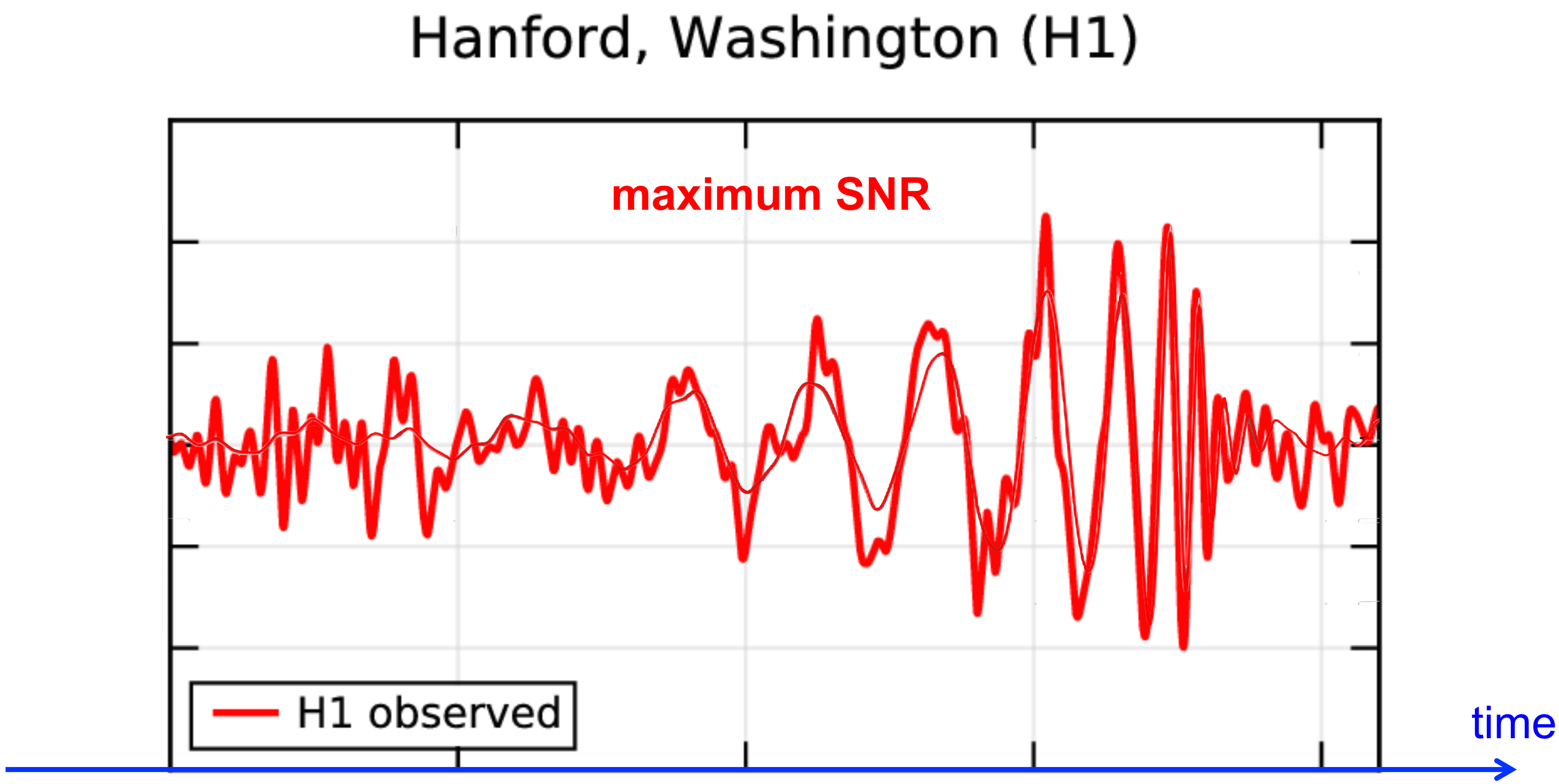
Matched filtering

- The SNR is calculated by taking a noise-weighted inner product

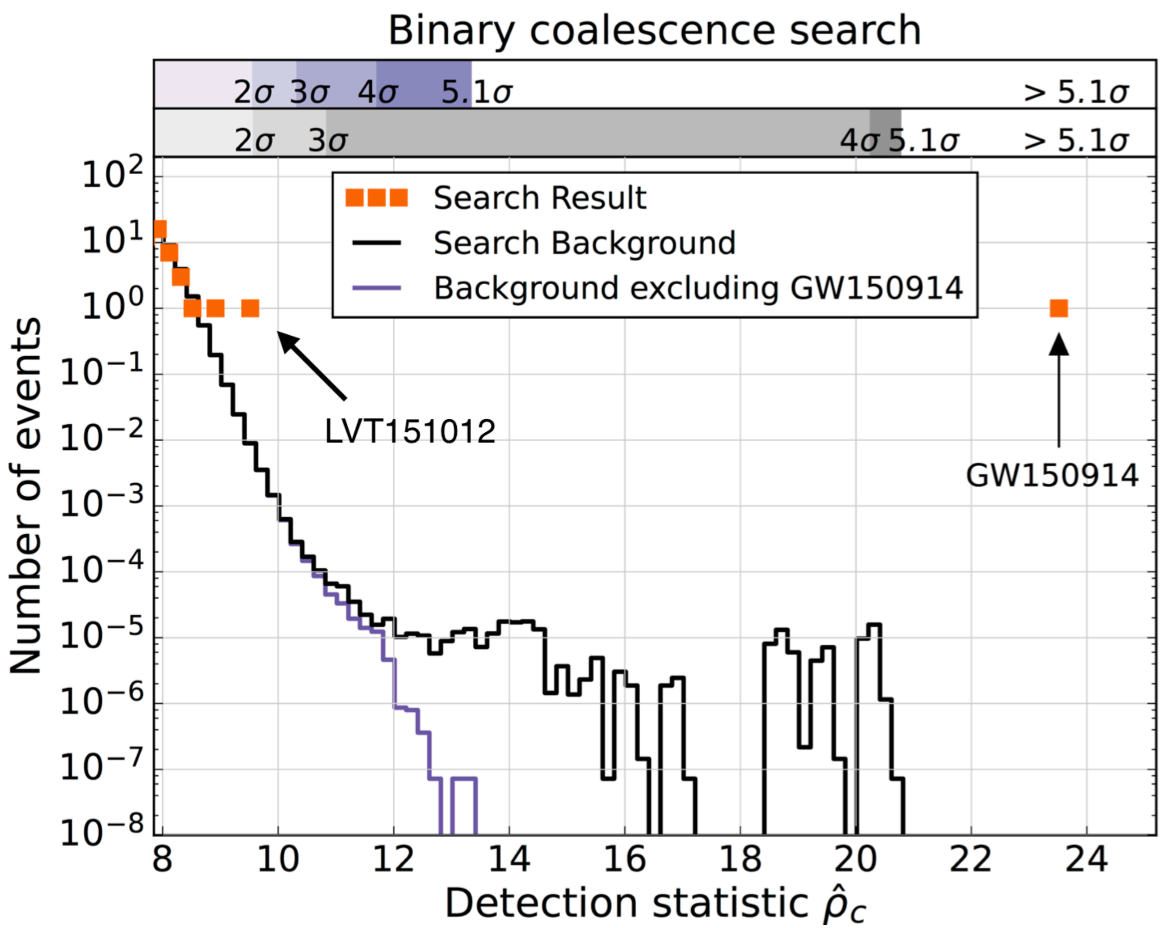
$$\rho^2 \sim \frac{\langle d|h\rangle^2}{\langle h|h\rangle}$$

where *d* is the data from the detectors and *h* is the signal template

- This is achieved in the time domain by convolving the template with the time series data



Estimating significance



The background is estimated by time-shifted data

GW150914 is louder than any background event and has an estimated significance of 1 in 203000 years

- **This is known as Bayes' theorem**

$$\underbrace{p(X|Y,I)}_{\text{posterior}} = \frac{\overbrace{p(Y|X,I)}^{\text{likelihood}} \times \overbrace{p(X|I)}^{\text{prior}}}{\underbrace{p(Y|I)}_{\text{evidence}}}$$

Thomas Bayes
1701-1761

- **For the purposes of statistical data analysis, we can interpret the above as**

$$p(parameter|data,I) \propto p(data|parameter,I) \times p(parameter|I)$$

- **Note that evidence is not used here and we will discuss this later**

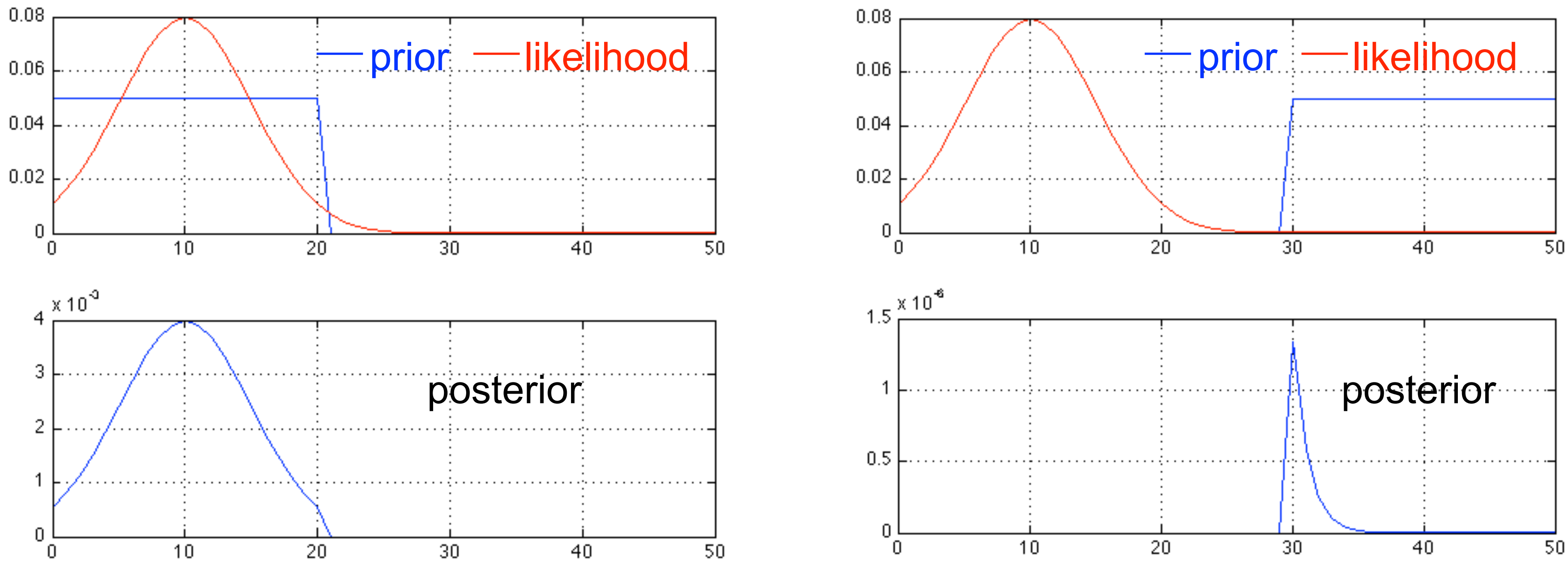
Bayesian inference

$$p(parameter|data, I) \propto p(data|parameter, I) \times p(parameter|I)$$

- **Here, the posterior probability represents our knowledge about the model given the data we have acquired**
- **The prior probability represents our state of knowledge about the model before any analysis of the data and this is modified by the acquisition of data through the likelihood function**

Revision: Bayesian inference

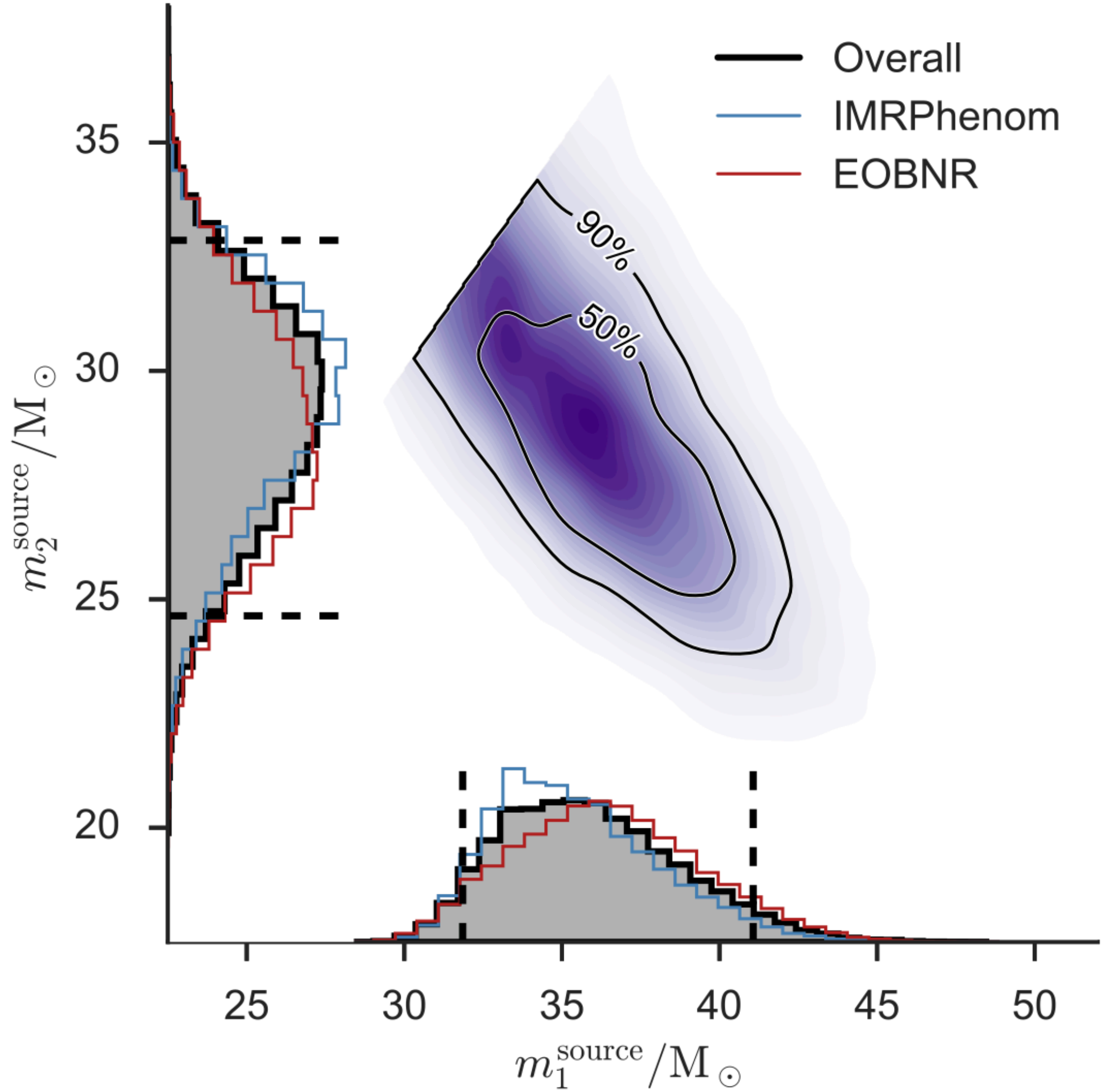
$$p(parameter|data, I) \propto p(data|parameter, I) \times p(parameter|I)$$



- **Inference about the model should be made using the posterior**
- **If we consider a broad, uniform prior, then the posterior is proportional to the likelihood**

Parameter estimation of GW150914

- The mass of the component black holes were estimated to be ~36 and 29 M_{sun}
- The post-merger black hole is estimated to be of 62 M_{sun}
-

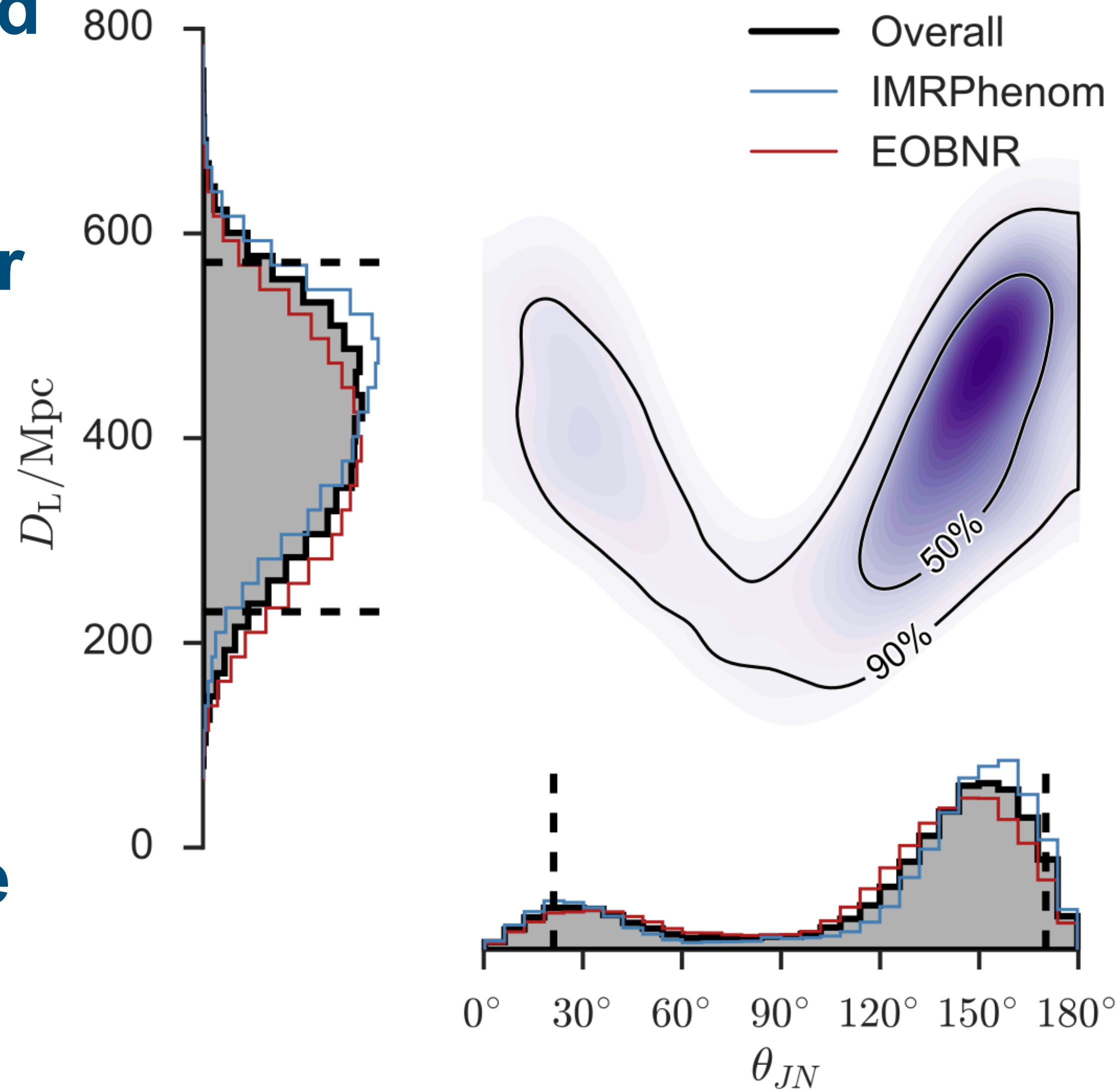


Parameter estimation of GW150914

- The measured amplitude is determined by 4 factors: mass, distance, orbital inclination and detector calibration

- The redshift of the source is z~0.1

- In principle, we can use galaxy catalogues to identify potential host galaxies



Bayesian model selection

- Recall Bayes' theorem where for desired model, M_A, and some observational data, D, we have

$$\underbrace{p(M_A|D,I)}_{\text{posterior}} = \frac{\overbrace{p(D|M_A,I)}^{\text{likelihood}} \times \overbrace{p(M_A|I)}^{\text{prior}}}{\underbrace{p(D|I)}_{\text{evidence}}}$$

- The posterior probability represents the state of our knowledge of the model ("the truth") in light of our observed data
- If we have a competing model or hypothesis, we use the ratio of the posterior probabilities for each model

$$\frac{p(M_A|D,I)}{p(M_B|D,I)} = \frac{p(M_A|I)}{p(M_B|I)} \times \frac{p(D|M_A,I)}{p(D|M_B,I)}$$

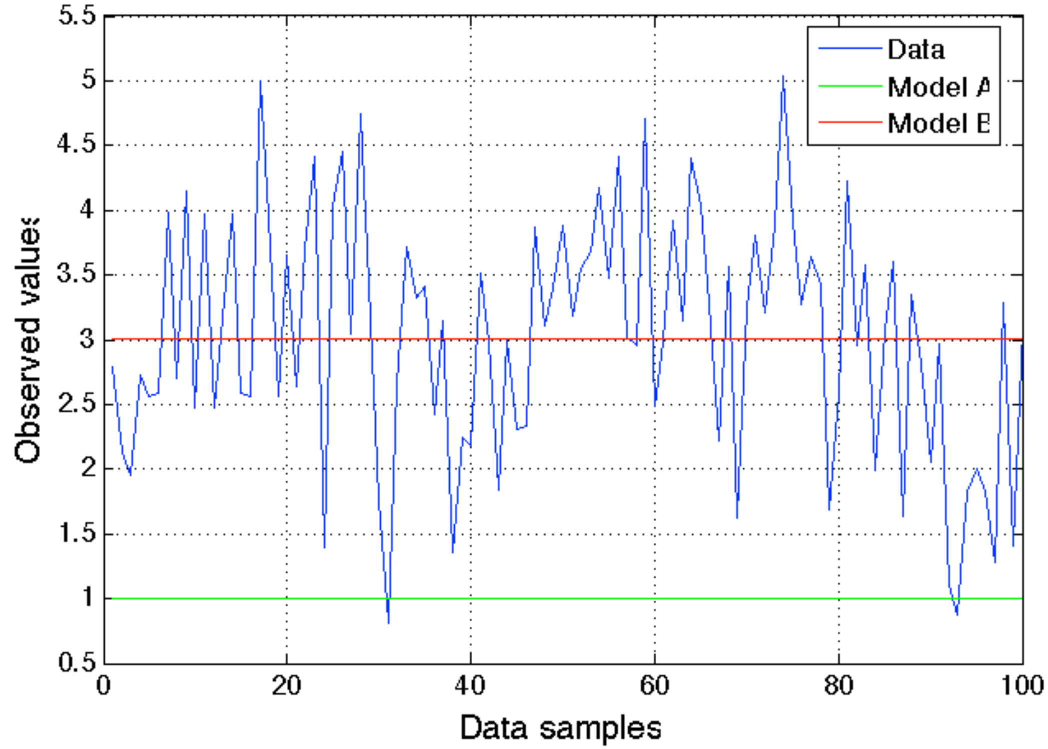
Bayesian model selection

- **This ratio of posterior probabilities is the odds ratio**

$$O_{AB} = \frac{p(M_A|D,I)}{p(M_B|D,I)} = \underbrace{\boxed{\frac{p(M_A|I)}{p(M_B|I)}}}_{\text{prior odds}} \times \underbrace{\boxed{\frac{p(D|M_A,I)}{p(D|M_B,I)}}}_{\text{Bayes factor}}$$

- **If $O_{AB} > 1$, M_A is preferred. If $O_{AB} < 1$, M_B is preferred**
- **If $O_{AB} = 1$, then there is insufficient information in the data to support either model**
- **There are a couple of points to note**
  - Prior odds: If we believe that one model is more likely than the other, then that bias is reflected in this ratio.
  - Bayes factor: The ratio of the likelihoods for each model, where we determine which model the data supports more

- **The prior odds is effectively just a number that quantifies our preference for one model over the other**
- **On the other hand, the Bayes factor reflects which model the data supports**
- **In the trivial example on the right, we compare the data (*n* = 100) to 2 single-valued models**
- **We assume a Gaussian likelihood with σ = 1 and calculate**



$$\frac{p(D|M_A, I)}{p(D|M_B, I)} = \prod_{i=1}^{n} \frac{e^{-(D_i-1)^2/2}}{e^{-(D_i-3)^2/2}} = 1.23 \times 10^{-89}$$

Model B is strongly favoured!

Marginalised likelihood

- The previous example was trivial since both models did not have any parameters
- We almost always encounter models with a certain number of parameters
- For such models, we obtain the likelihoods by marginalising over all desired model parameters
- Marginalisation:

$$p(X|I) = \int_{-\infty}^{\infty} p(X,Y|I)dY = \int_{-\infty}^{\infty} p(X|Y,I)p(Y|I)dY$$

- Note that marginalisation calculates the average likelihood over all parameter values, weighted by the prior for that parameter

Marginalised likelihood

- For a model, *M*, with parameters θ, we obtain the likelihood by

$$p(D|M,I) = \int_{-\infty}^{\infty} p(D|\theta,M,I)p(\theta|M,I)d\theta$$

- For more parameters, you introduce an integral for each parameter to marginalise over
- This integral can be evaluated analytically though this integral is often evaluated numerically
- The prior range can restrict the limits of the integration

Occam factor

- Lets again consider two models, M_A and M_B
- Both models have a Gaussian likelihood
- M_A does not have any parameters while M_B has a single parameter, θ

$$p(D|M_B, I) = \int_{-\infty}^{\infty} p(D|\theta, M_B, I)p(\theta|M_B, I)d\theta$$

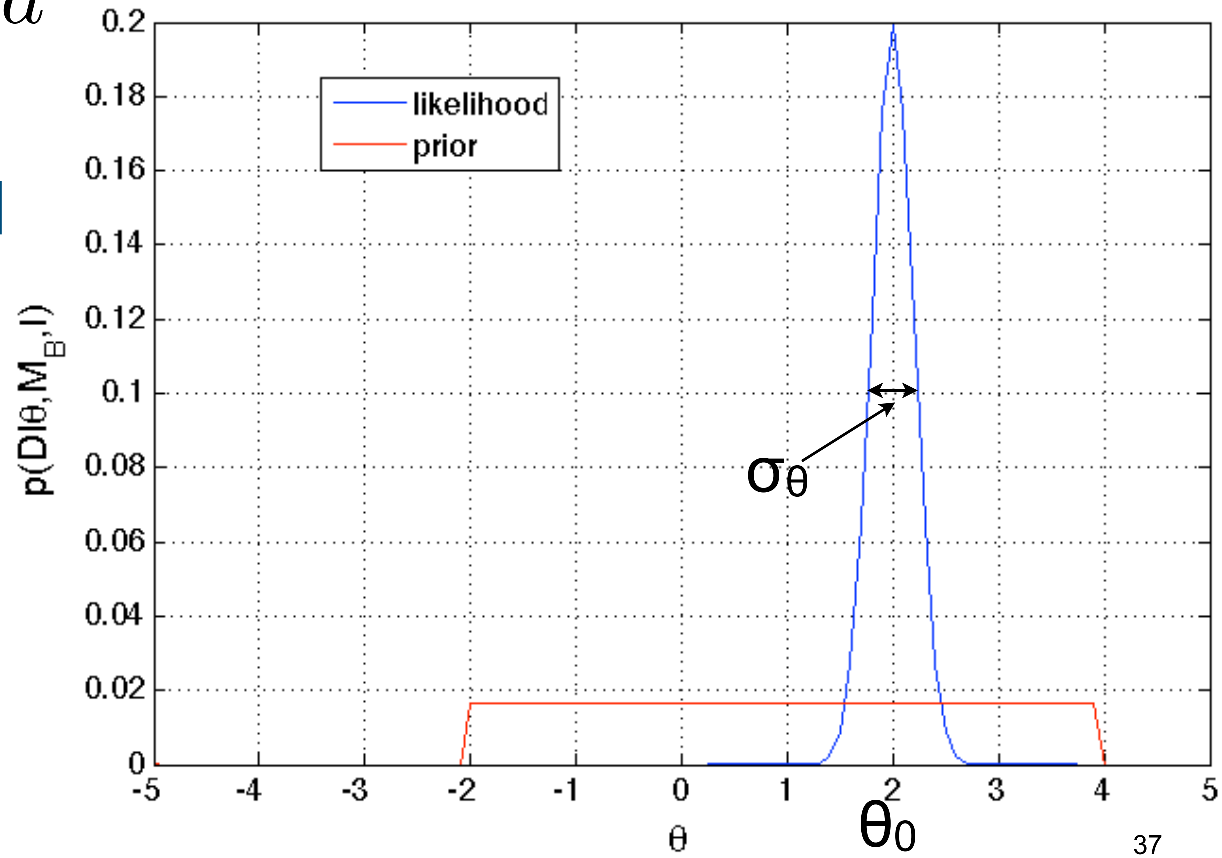
- The parameter θ has a uniform prior between the values *a* and *b*, so

$$p(D|M_B, I) = \frac{1}{b-a}\int_{a}^{b} p(D|\theta, M_B, I)d\theta$$

- If we have *a* and *b* wide enough that it does not significantly truncate the Gaussian likelihood, then the likelihood for M_B becomes

$$p(D|M_B, I) = \frac{\sigma_\theta \sqrt{2\pi}}{b-a} p(D|\theta_0, M_B, I)$$

- where θ_0 is the peak of the likelihood and σ_θ is its standard deviation



Occam factor

- So, our odds ratio is

$$O_{AB} = \frac{p(M_A|I)}{p(M_B|I)} \times \frac{p(D|M_A, I)}{p(D|M_B, I)}$$

$$= \frac{p(M_A|I)}{p(M_B|I)} \times \frac{p(D|M_A, I)}{p(D|\theta_0, M_B, I)} \times \frac{b-a}{\sigma_\theta \sqrt{2\pi}}$$

- We previously required that $b - a > \sigma_\theta$, so the factor on the right is greater than 1
- Such a factor (often referred to as Occam's factor) penalises M_B since it has more degrees of freedom to fit the data than M_A