

Data analysis for transient gravitational waves

Gravitational wave mini-school
at
Beijing Normal University
15-18 September, 2016

Heng Ik Siong (王毅雄)



University
of Glasgow

- **Introduction**

- Normal, Student-t and χ^2 distributions
- Power spectral density

- **Bayesian inference**

- parameter estimation

- **Bayesian model selection**

- **Methods for identifying gravitational wave transients**

- generic transients
- matched filtering

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

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

- 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

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

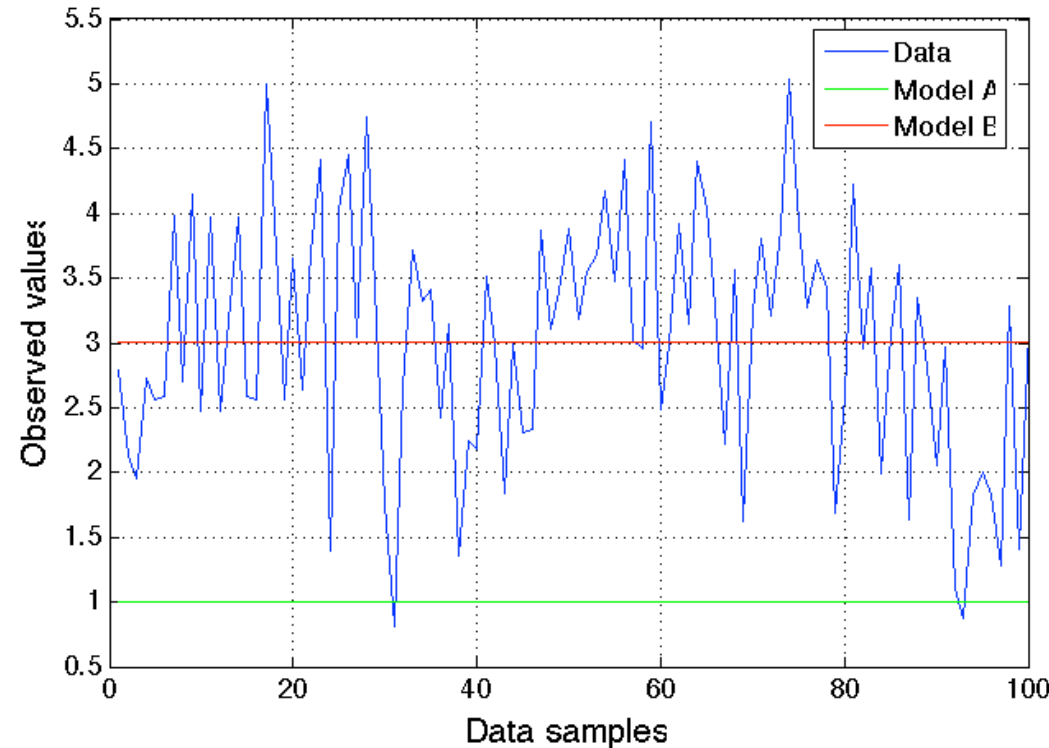
- This ratio of posterior probabilities is the odds ratio

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

- 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 $\sigma = 1$ 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!

- 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

- 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

- 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

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

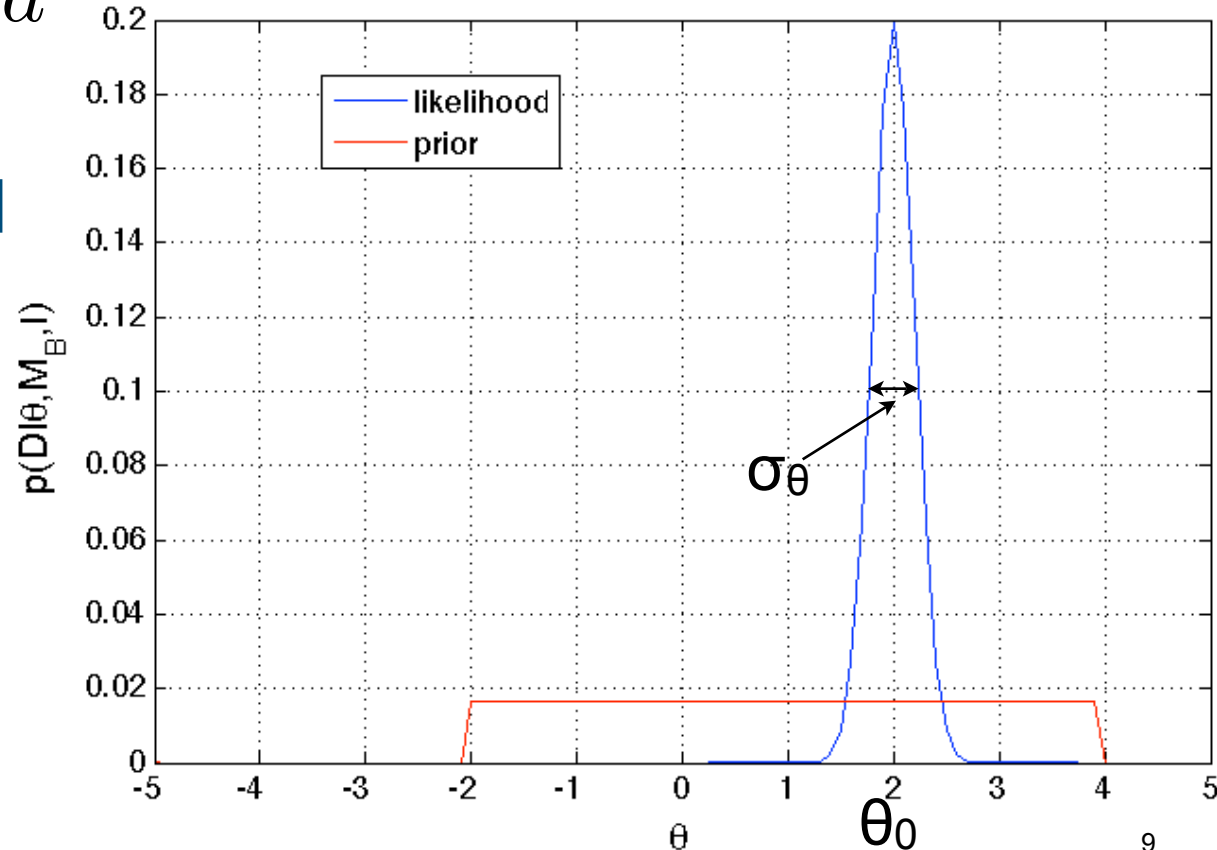
- The parameter θ has a uniform prior between the values a and b , so

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

- 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



- So, our odds ratio is

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

- 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

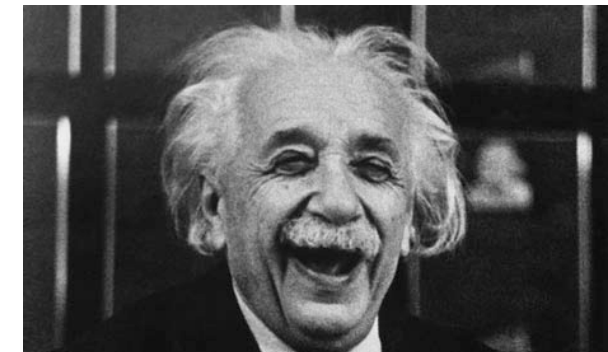
- For parameters with wide uniform priors, we can approximate the likelihood by

$$p(D|M, I) \approx L_{\max} \frac{\text{likelihood range}}{\text{prior range}}$$

- Note that we want the prior range for our parameters to be as close to the likelihood range
- If we choose to have a broad range for our parameters (broad prior), then we are penalised for being cautious
- “Everything must be made as simple as possible, but not simpler”



William of Ockham
1288 to ~1348



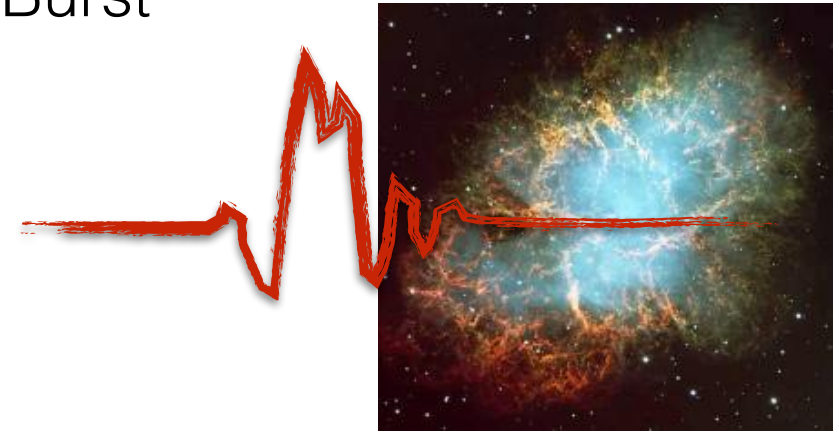
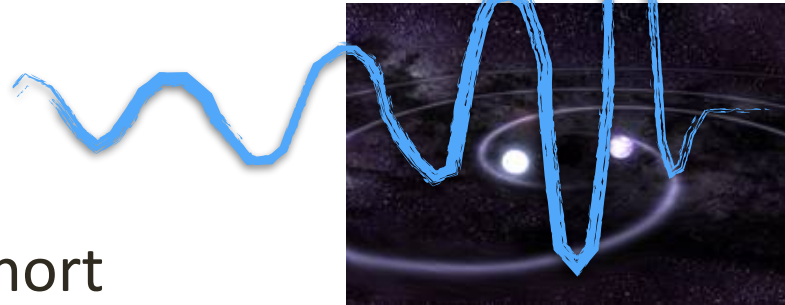
Gravitational wave sources types

modelled

unmodelled

Compact Binary
Coalescence

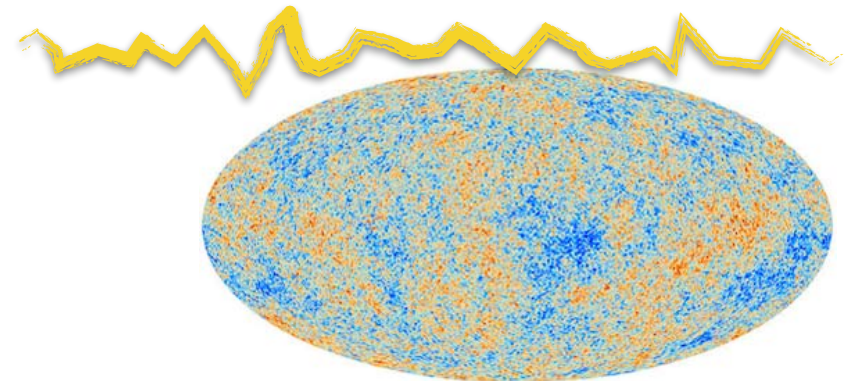
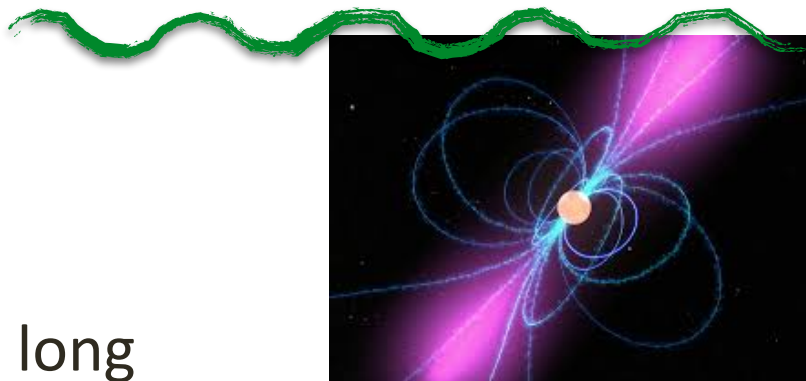
Burst



short

Continuous

Stochastic



long

- **Two kinds of burst and CBC searches**

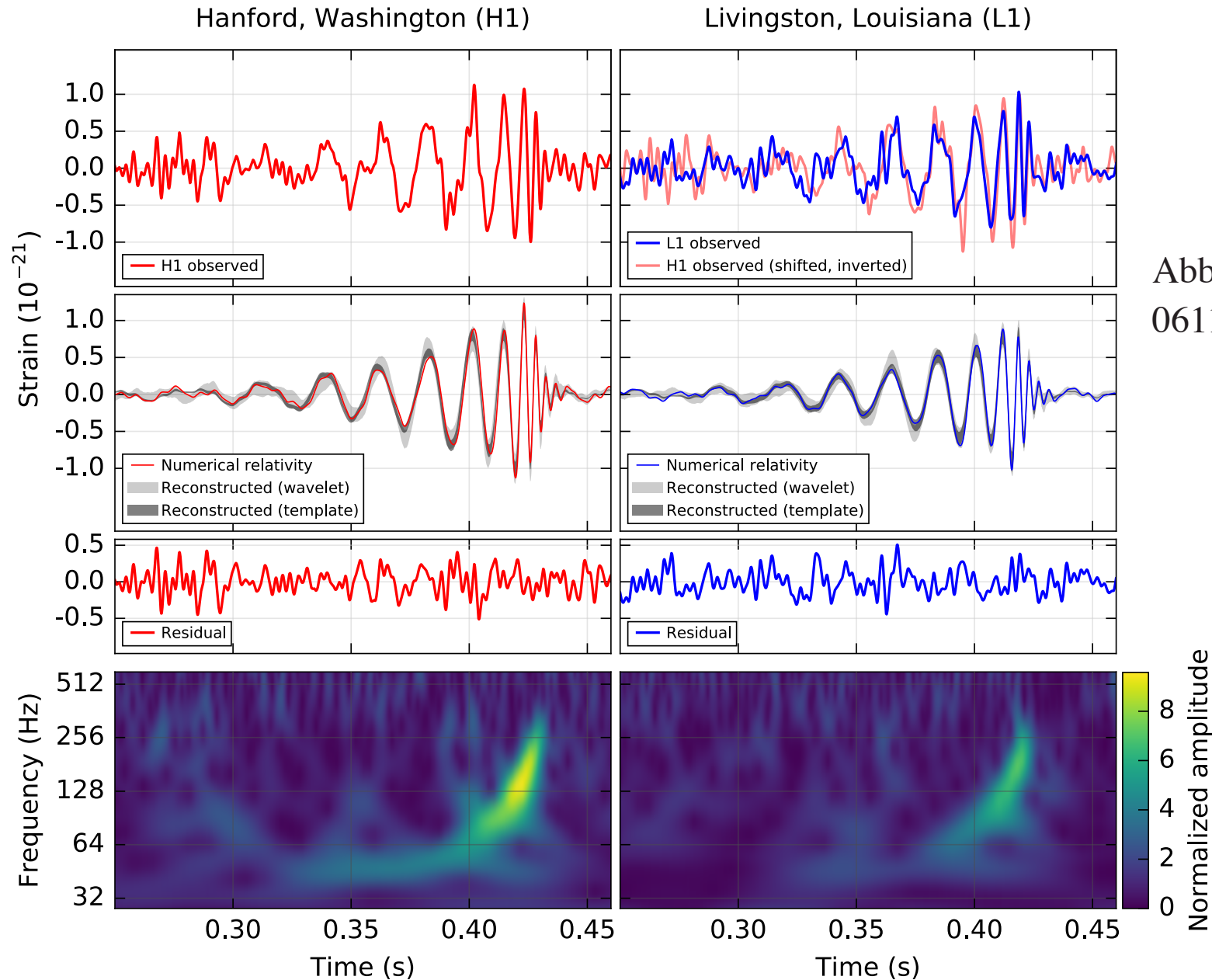
- all-sky search: maximise detection efficiency across the whole sky
- triggered search: search triggered by non-GW observation (eg. GRBs)

- **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

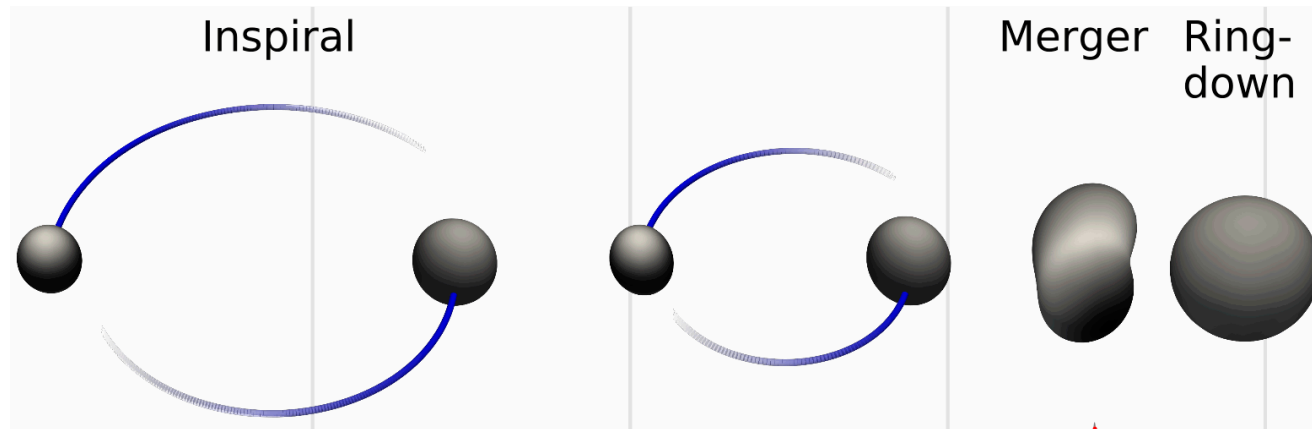


Abbot et al., PRL 116,
061102 (2016)

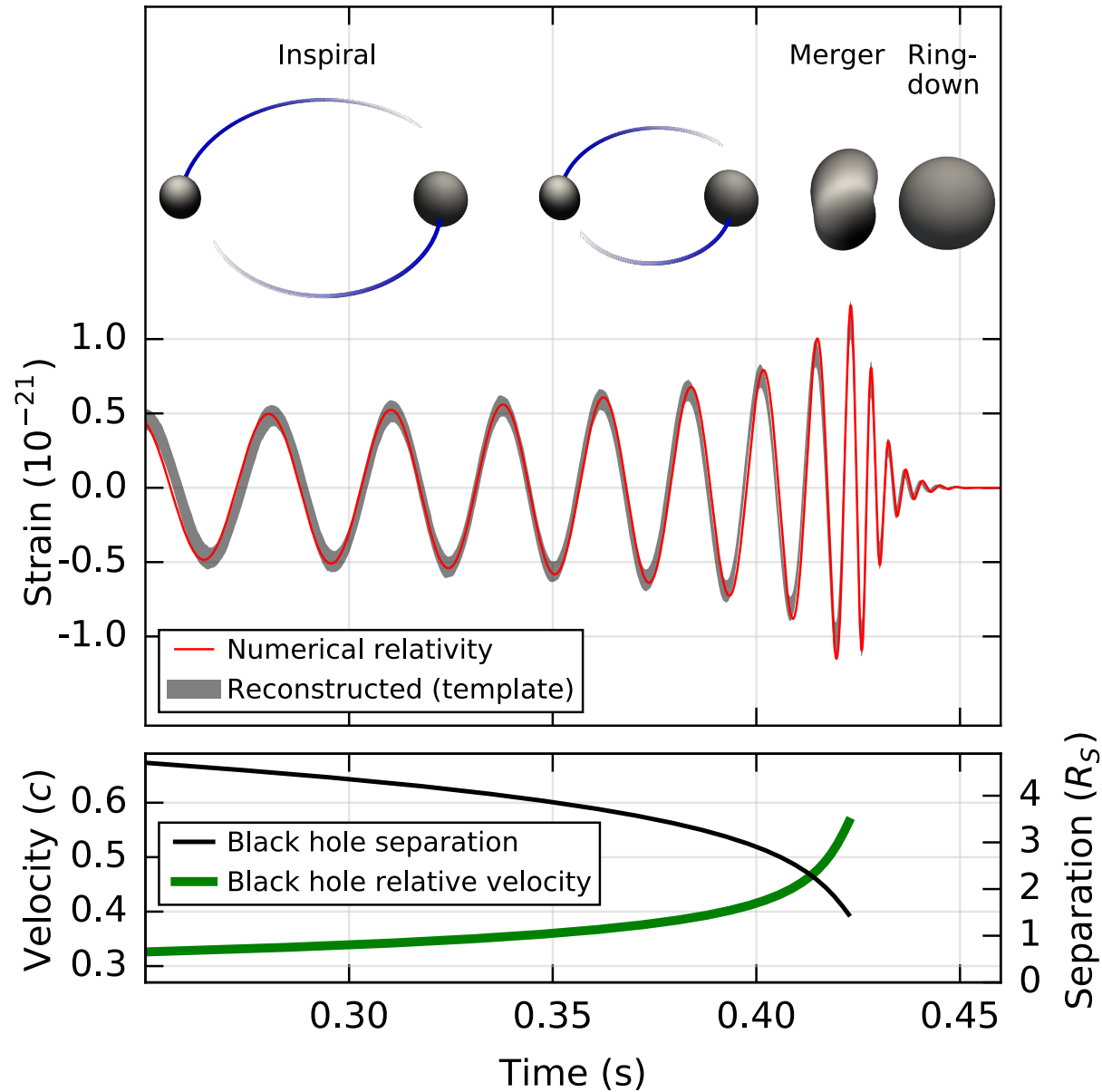
Detection by generic transient searches

- 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].”



Evolution of the signal



- Links to papers

<https://papers.ligo.org/>

- LIGO Open Science Center (LOSC): download 30 sec of data around the event plus more!

<http://losc.ligo.org/>

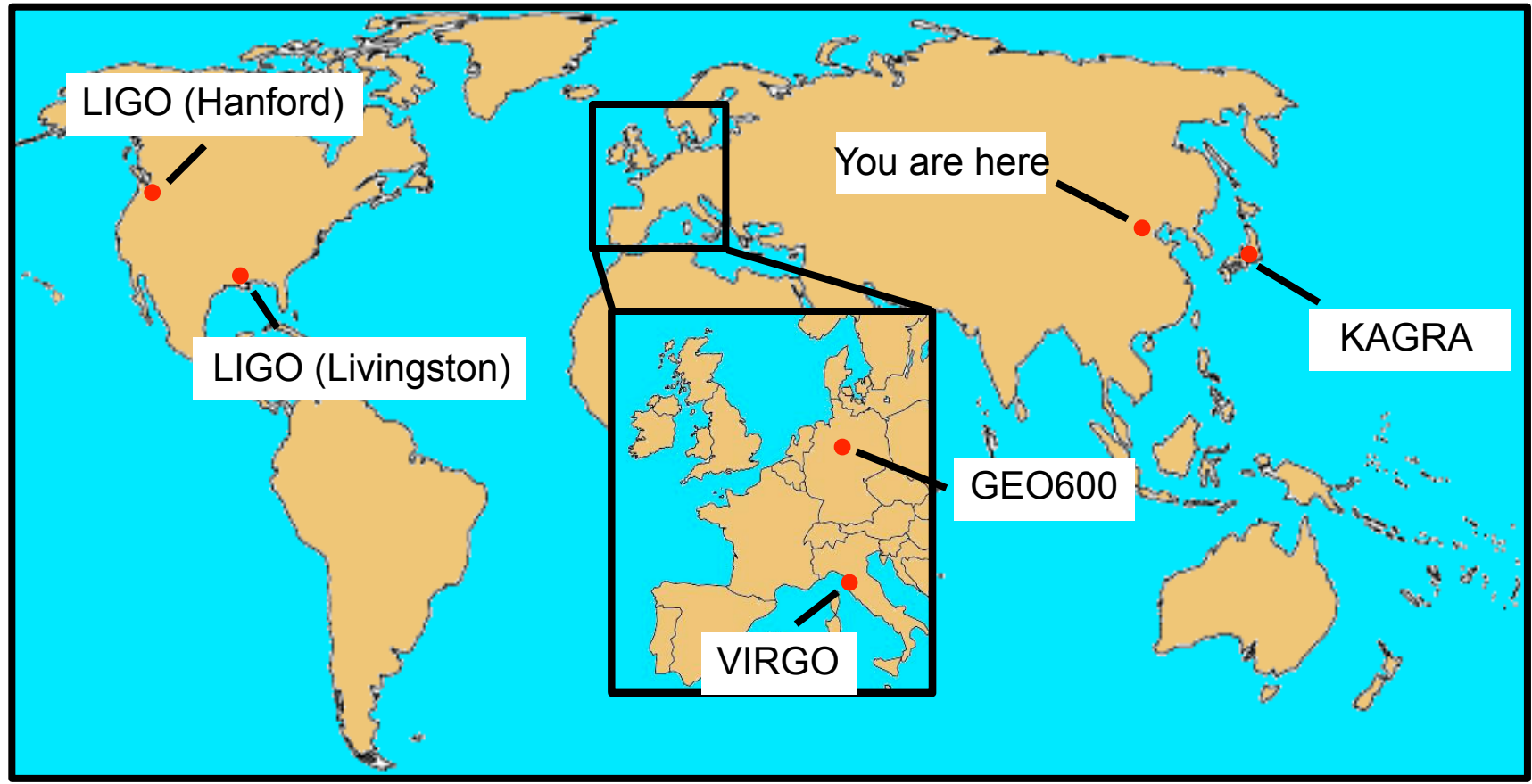
- Science summary of detection

https://www.ligo.caltech.edu/system/media_files/binaries/301/original/detection-science-summary.pdf

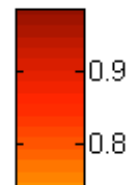
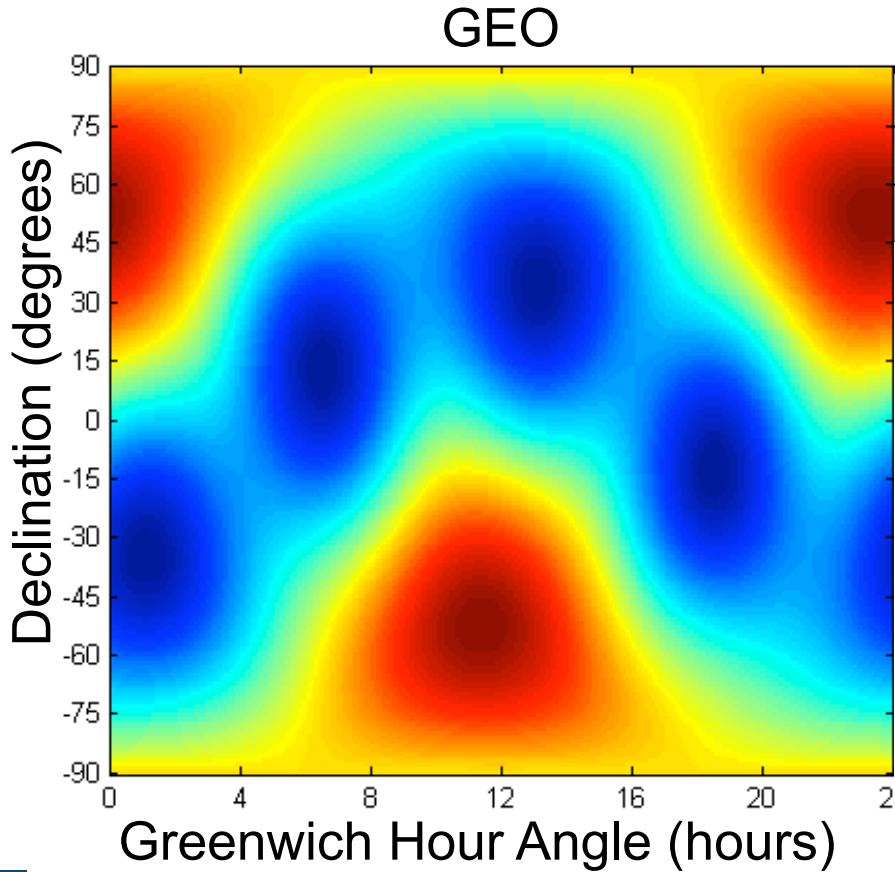
- Many videos, images and other resources available from

<https://www.ligo.caltech.edu/detection>

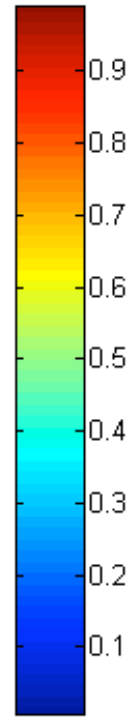
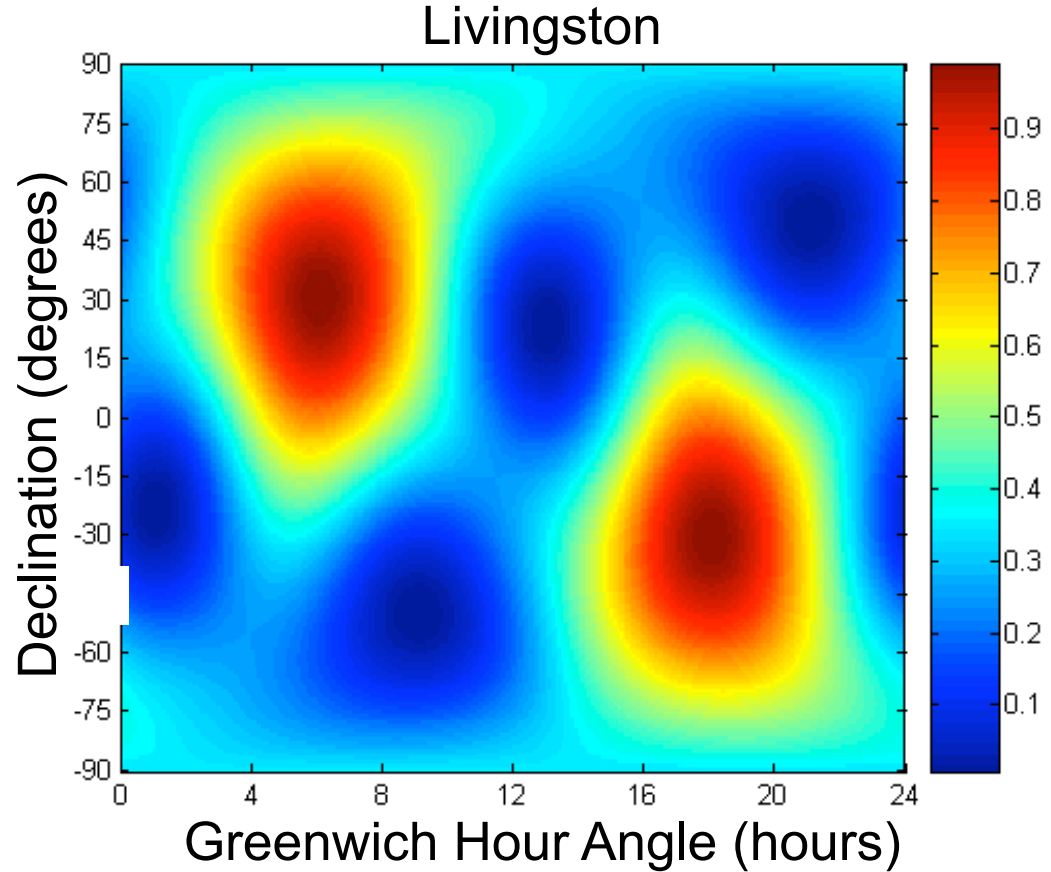
Interferometric detectors worldwide



Directional sensitivity



$$F_+^2 + F_x^2$$



- Lets formulate our data such that

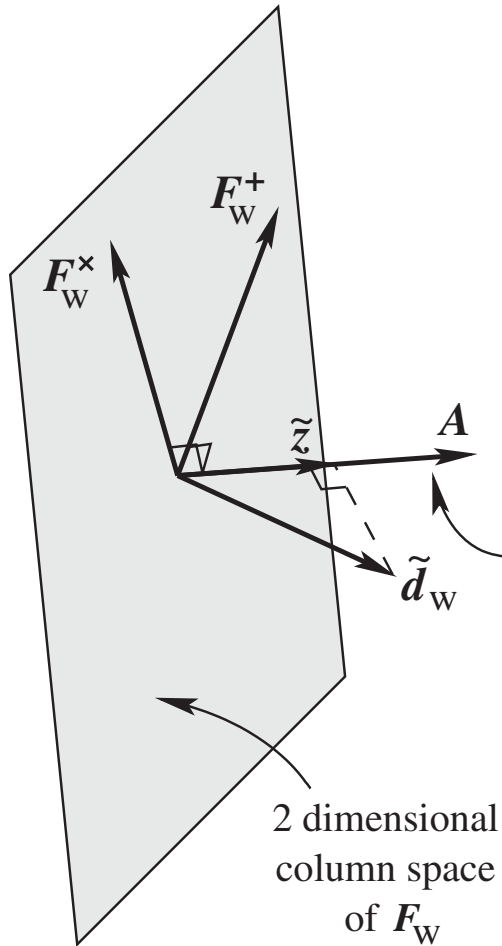
$$\tilde{\mathbf{d}}_w = \mathbf{F}_w \tilde{\mathbf{h}} + \tilde{\mathbf{n}}_w,$$

or

$$\begin{array}{c} \text{data} \\ \left[\begin{array}{c} \tilde{d}_{w1} \\ \tilde{d}_{w2} \\ \vdots \\ \tilde{d}_{wD} \end{array} \right] \end{array} = \begin{array}{c} \text{antenna pattern} \\ \left[\begin{array}{cc} F_{w1}^+ & F_{w1}^\times \\ F_{w2}^+ & F_{w2}^\times \\ \vdots & \vdots \\ F_{wD}^+ & F_{wD}^\times \end{array} \right] \end{array} \begin{array}{c} \text{signal} \\ \left[\begin{array}{c} \tilde{h}_+ \\ \tilde{h}_\times \end{array} \right] \end{array} + \begin{array}{c} \text{noise} \\ \left[\begin{array}{c} \tilde{n}_{w1} \\ \tilde{n}_{w2} \\ \vdots \\ \tilde{n}_{wD} \end{array} \right], \end{array}$$

$$\begin{array}{c} \text{sky location} \\ \swarrow \\ \mathbf{F}_w(\hat{\Omega}_s, f) \end{array} \begin{array}{c} \text{frequency} \\ \swarrow \\ f \end{array} \equiv \left[\mathbf{F}_w^+ \quad \mathbf{F}_w^\times \right] = \left[\begin{array}{cc} F_{w1}^+ & F_{w1}^\times \\ F_{w2}^+ & F_{w2}^\times \\ \vdots & \vdots \\ F_{wD}^+ & F_{wD}^\times \end{array} \right]$$

Chatterji et al, PRD 74, 082005 (2006)



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

$$A F_w = 0.$$

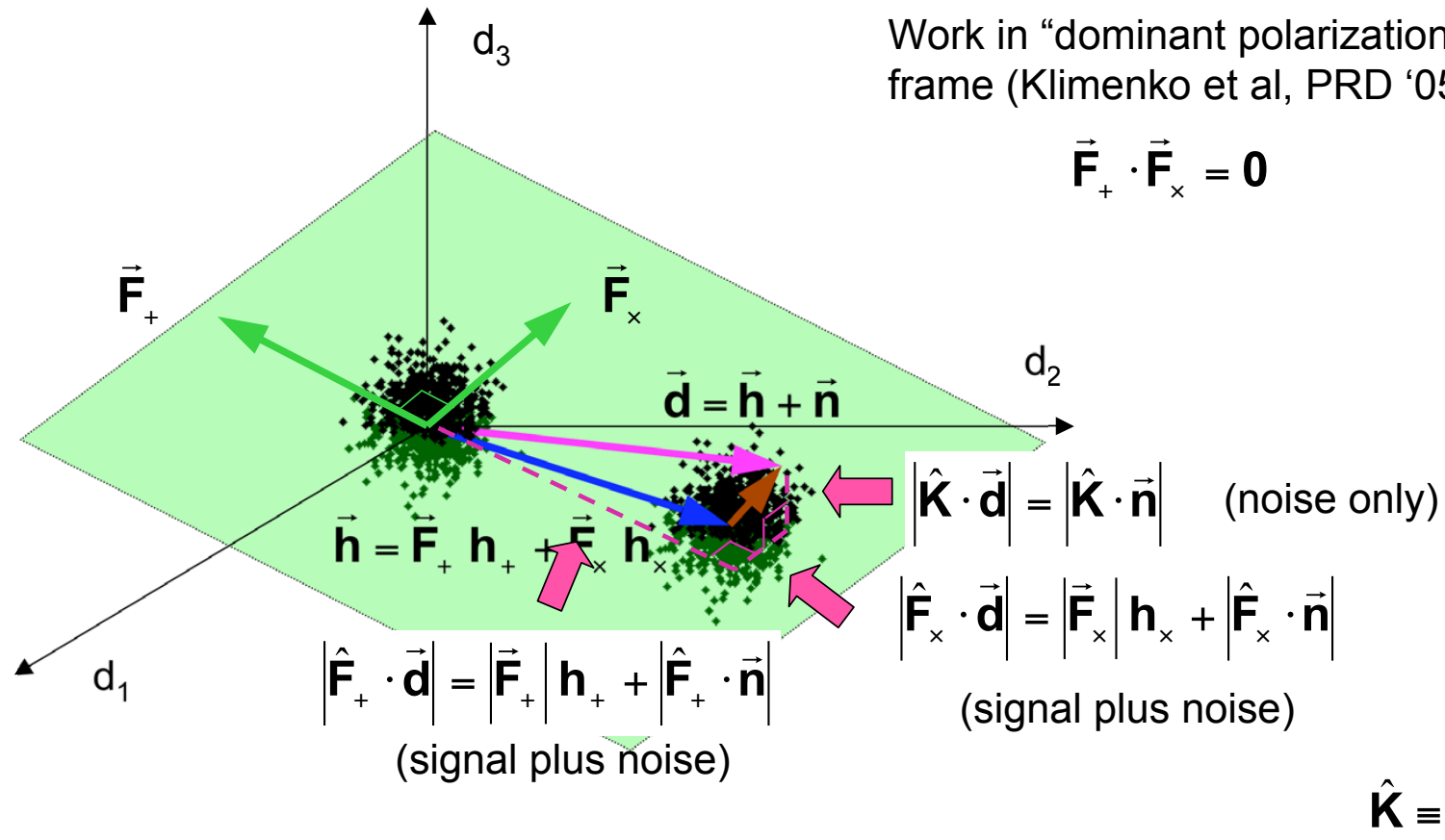
- then, we can use it to construct a null stream

$$\tilde{z} \equiv A \tilde{d}_w = A F_w \tilde{h} + A \tilde{n}_w = A \tilde{n}_w.$$

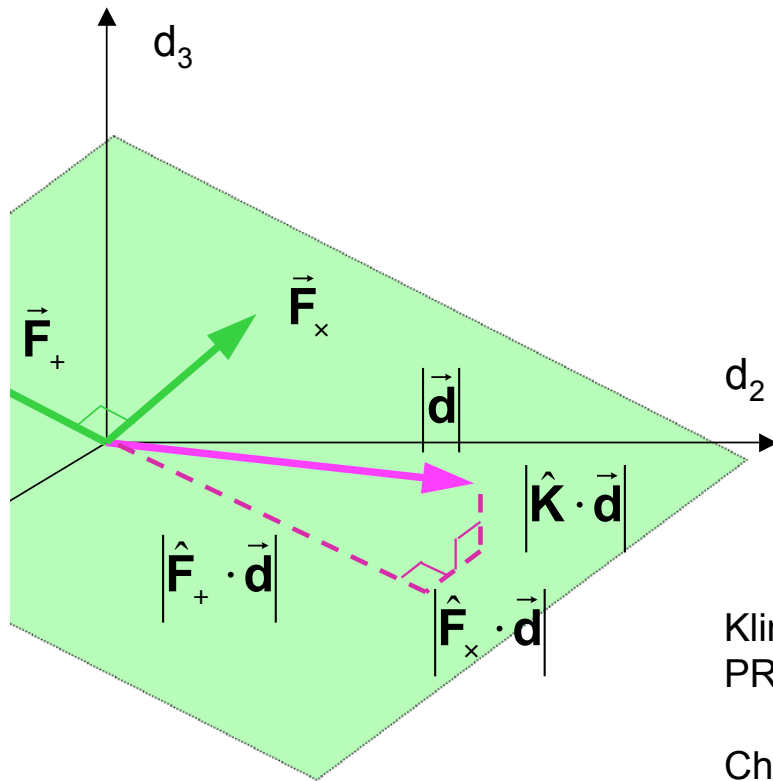
- For 3 detectors, A is constructed by

$$A = \frac{F_w^+ \times F_w^x}{|F_w^+ \times F_w^x|}$$

Geometric View: One TF Pixel



Likelihood or Energy Measures



Likelihoods used in this analysis:

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

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

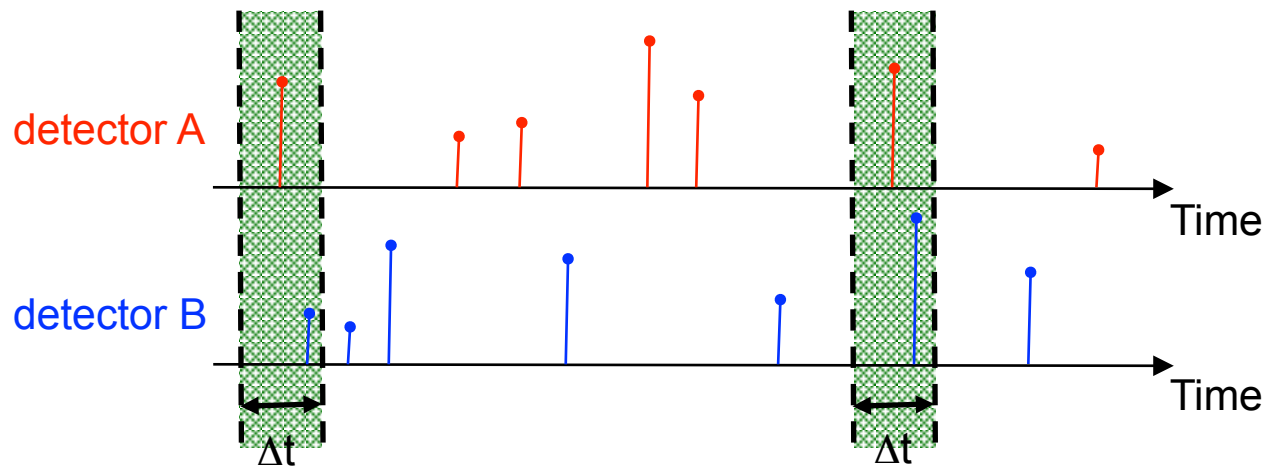
Incoherent Energy $E_{\text{INC}} = \sum_{\alpha} |\mathbf{K}_{\alpha} \cdot \mathbf{d}_{\alpha}|^2$

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)

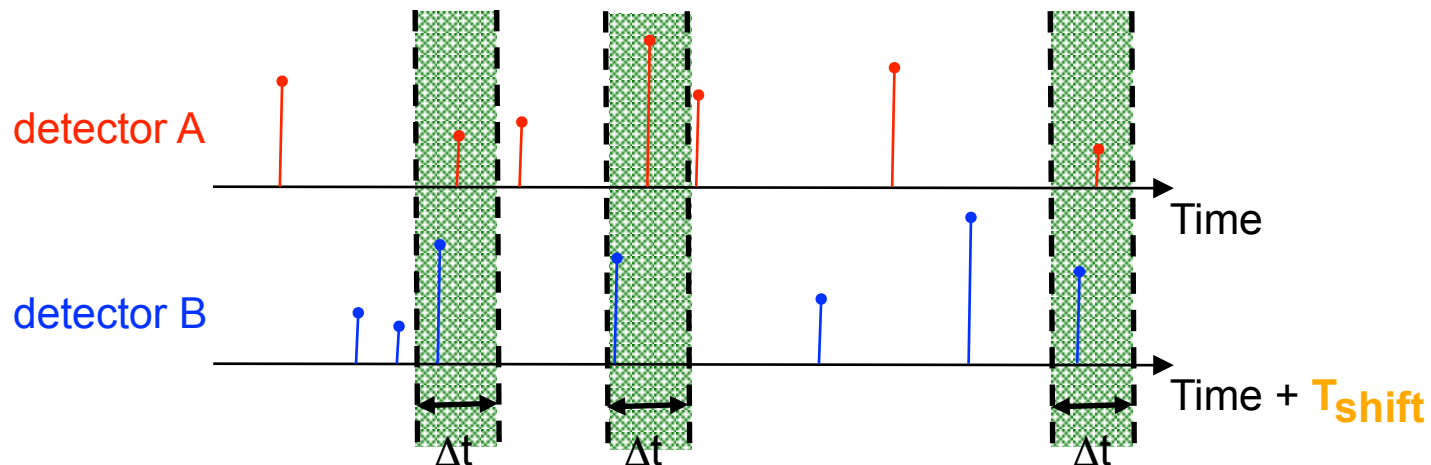
Coincidence analysis

- **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

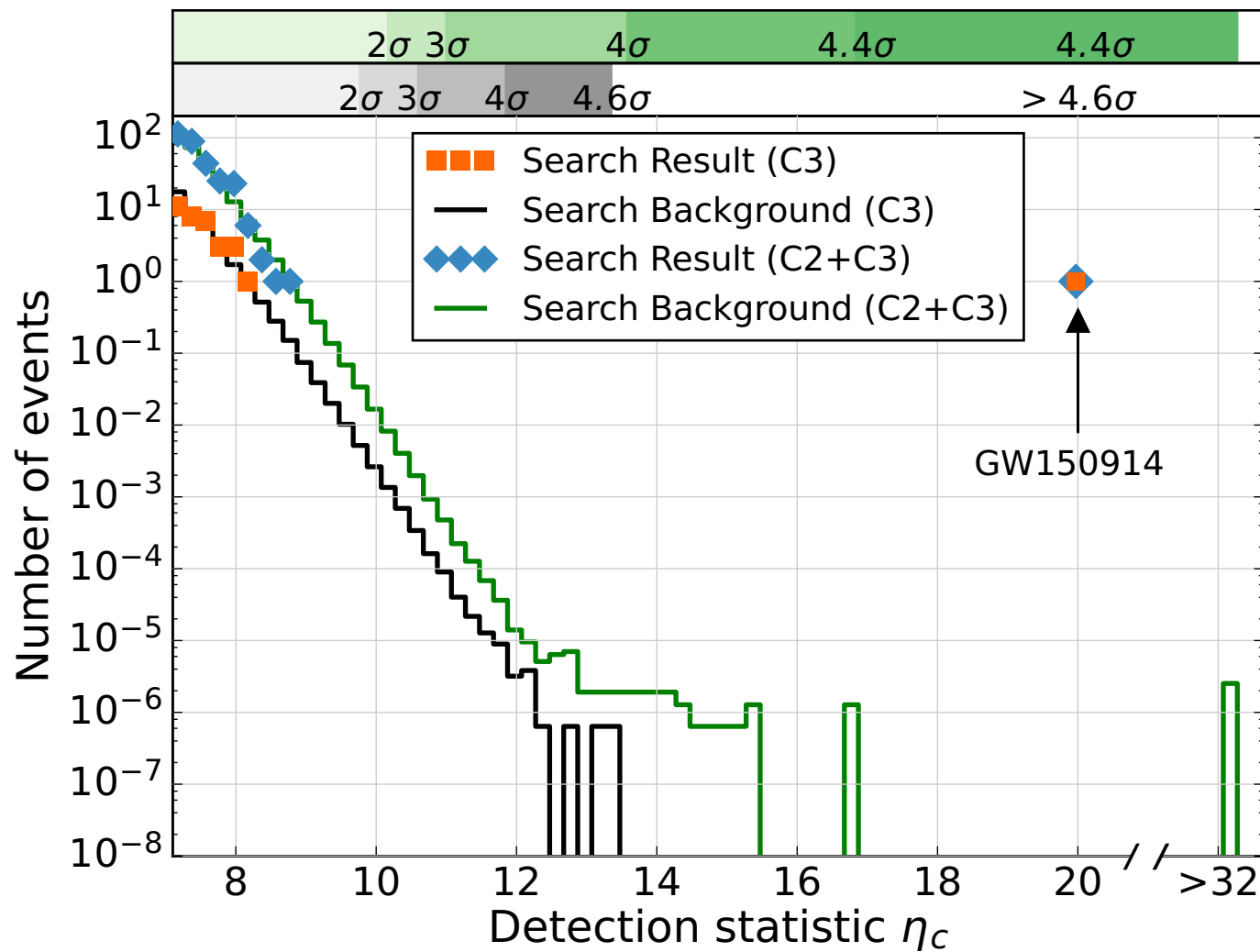


Coincidence analysis

- **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

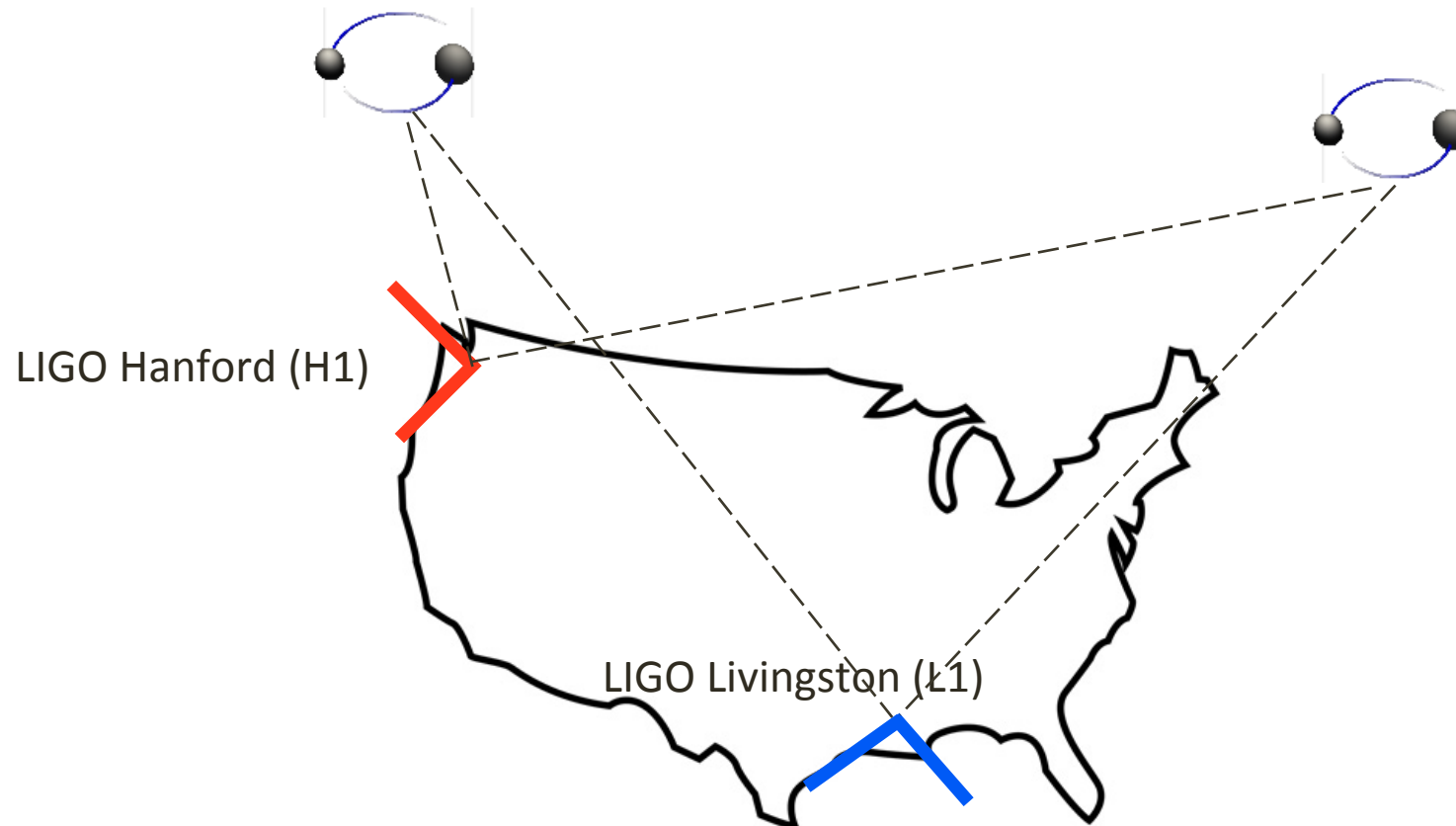


Generic transient search

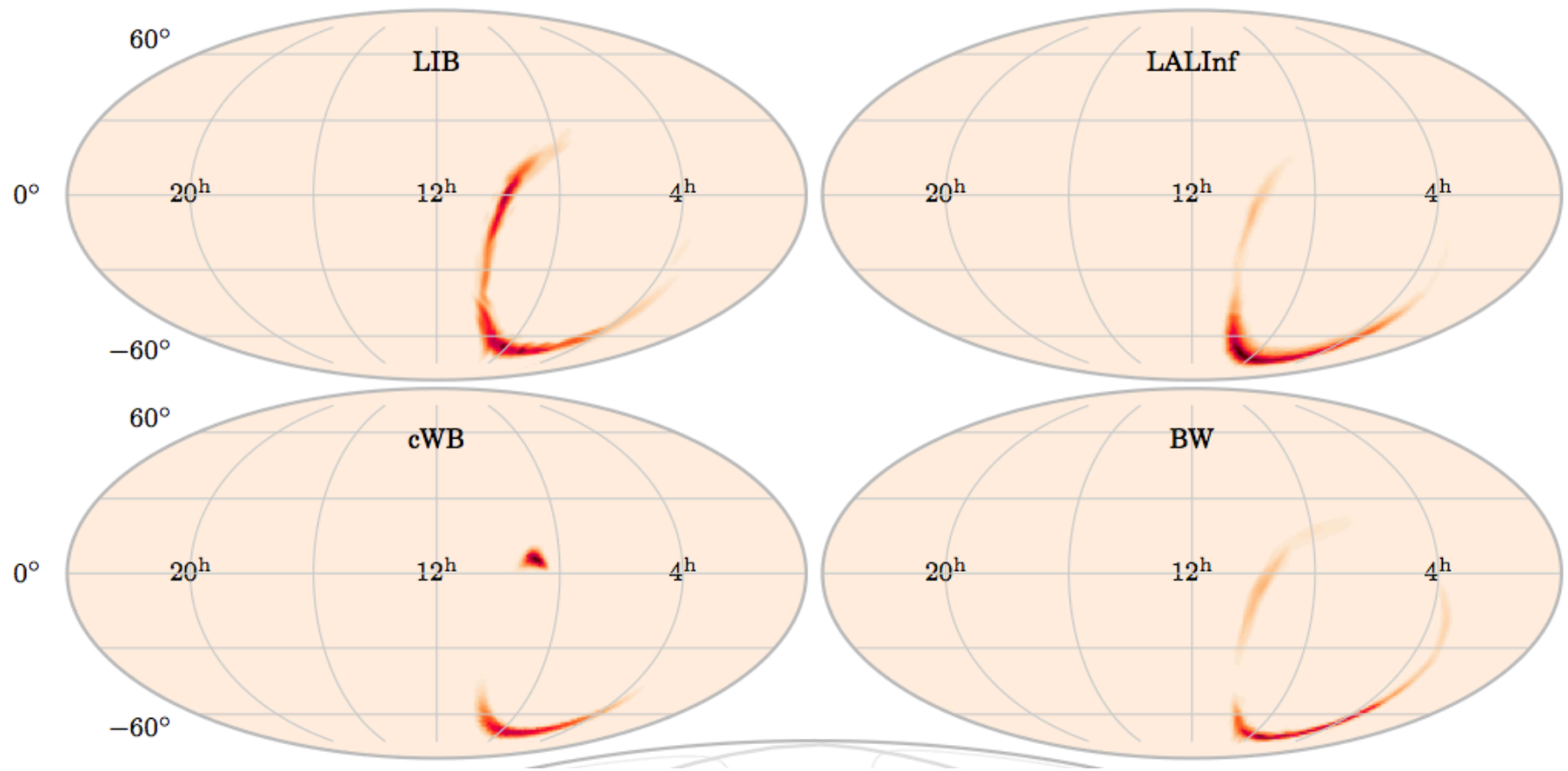


Rapid sky localisation

- 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

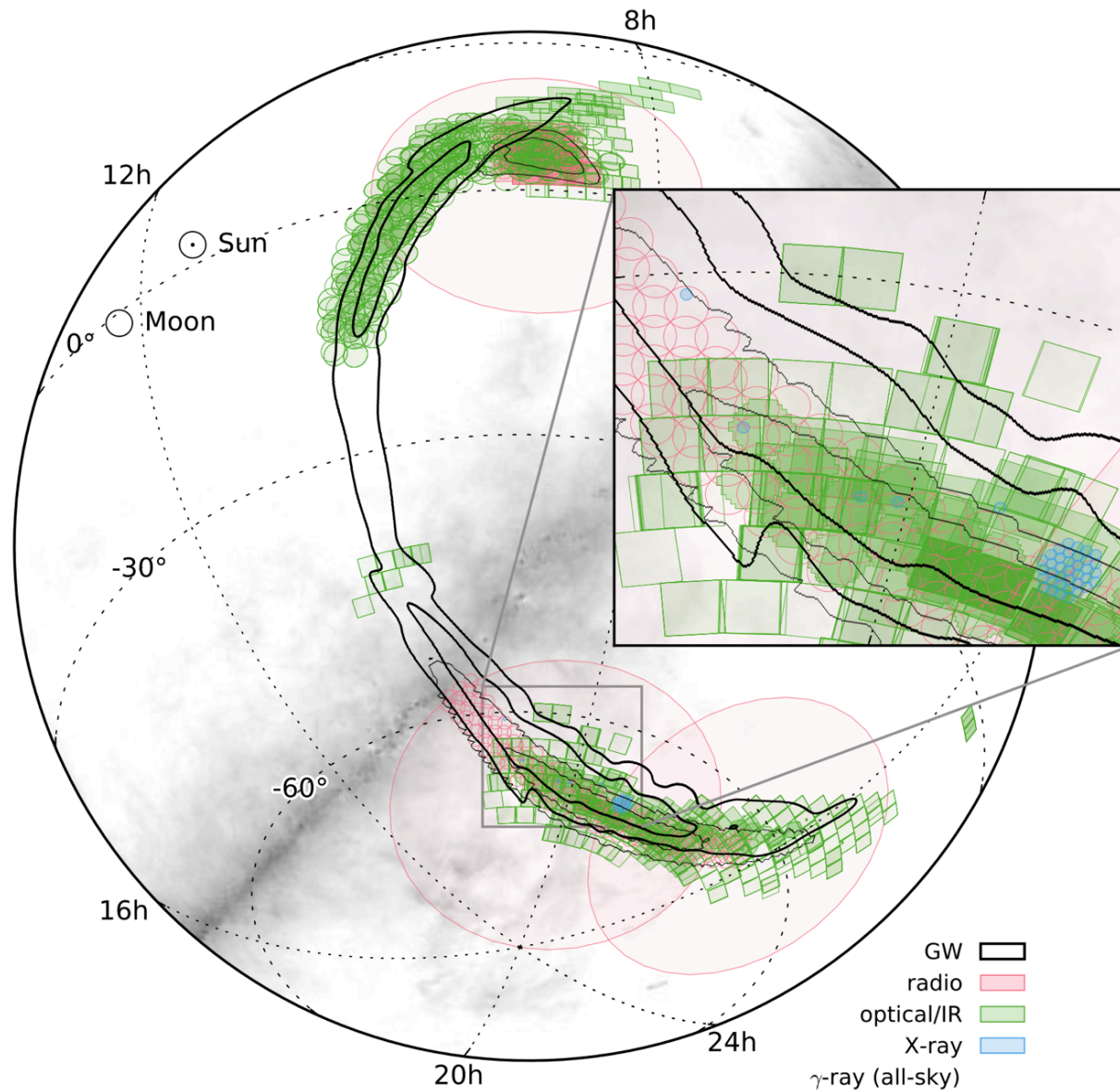


Rapid sky localisation



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

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**

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

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

- **The observed signal is a function of the antenna patterns**

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

- **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 noise**

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

- **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:

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

- 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 $\rho^2 = \frac{\Sigma^2}{N^2}$

- 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

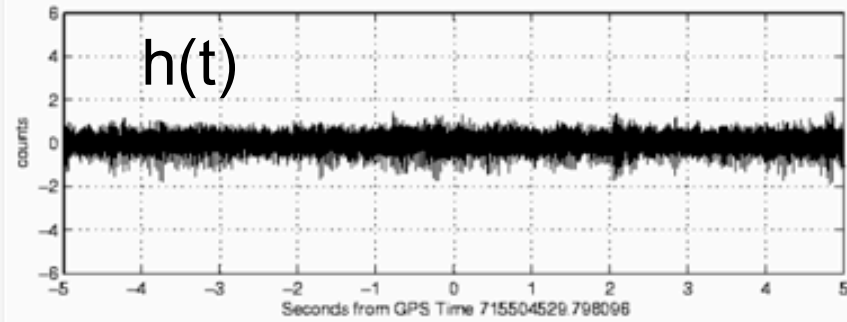
$$\tilde{q}(f) \propto \frac{\tilde{h}(f) e^{2\pi i f(\tau - t_a)}}{S_h(f)}$$

- Substituting this, we find

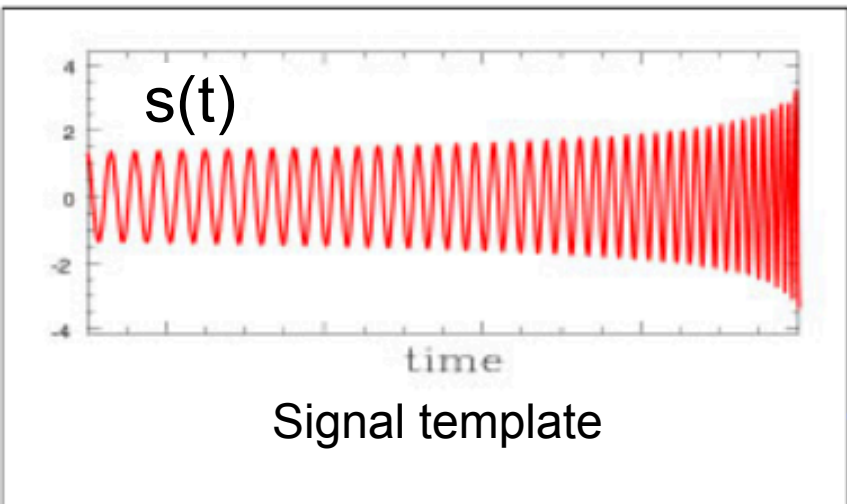
$$\rho_{\text{opt}}^2 = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_h(f)} df$$

Calculating Inspiral SNR

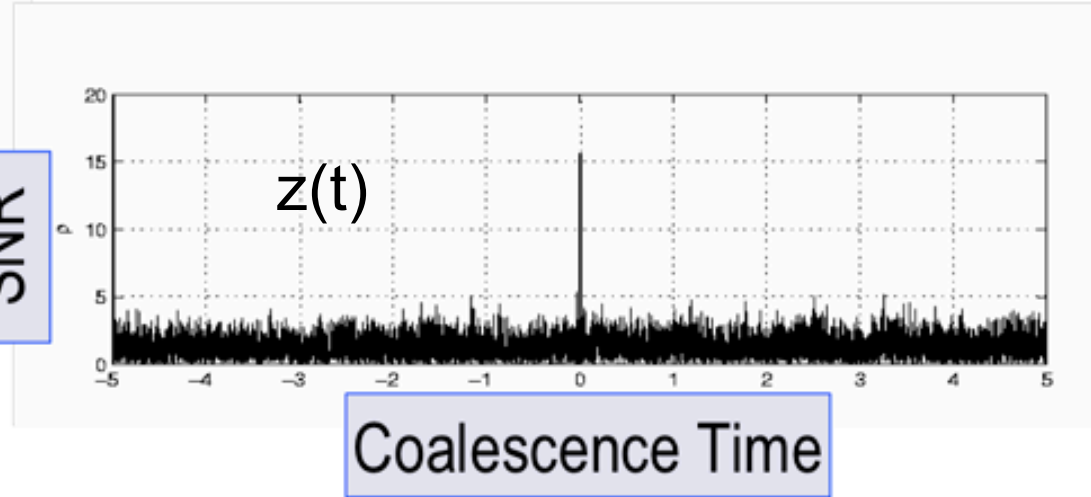
GW Channel
+ simulated inspiral



(X)



SNR



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

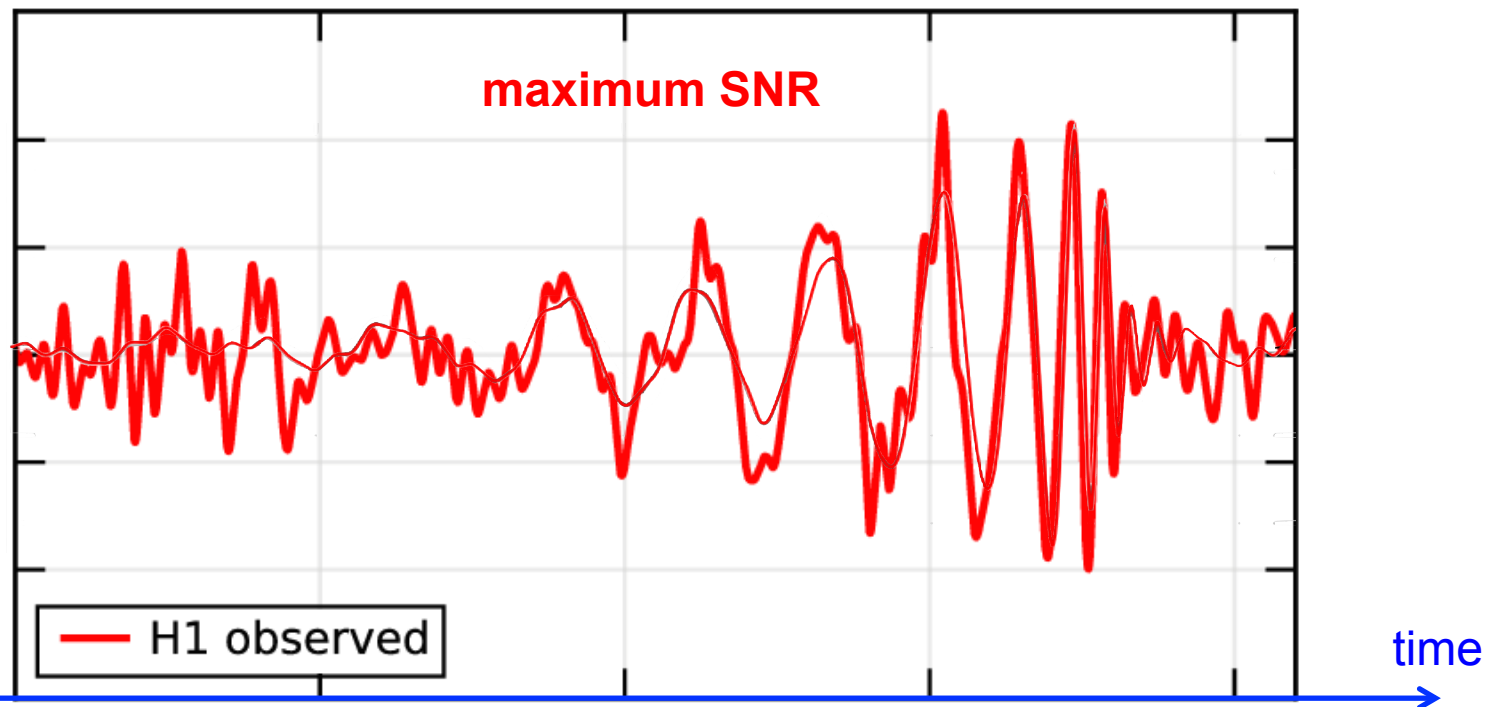
- The SNR is calculated by taking a noise-weighted inner product

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

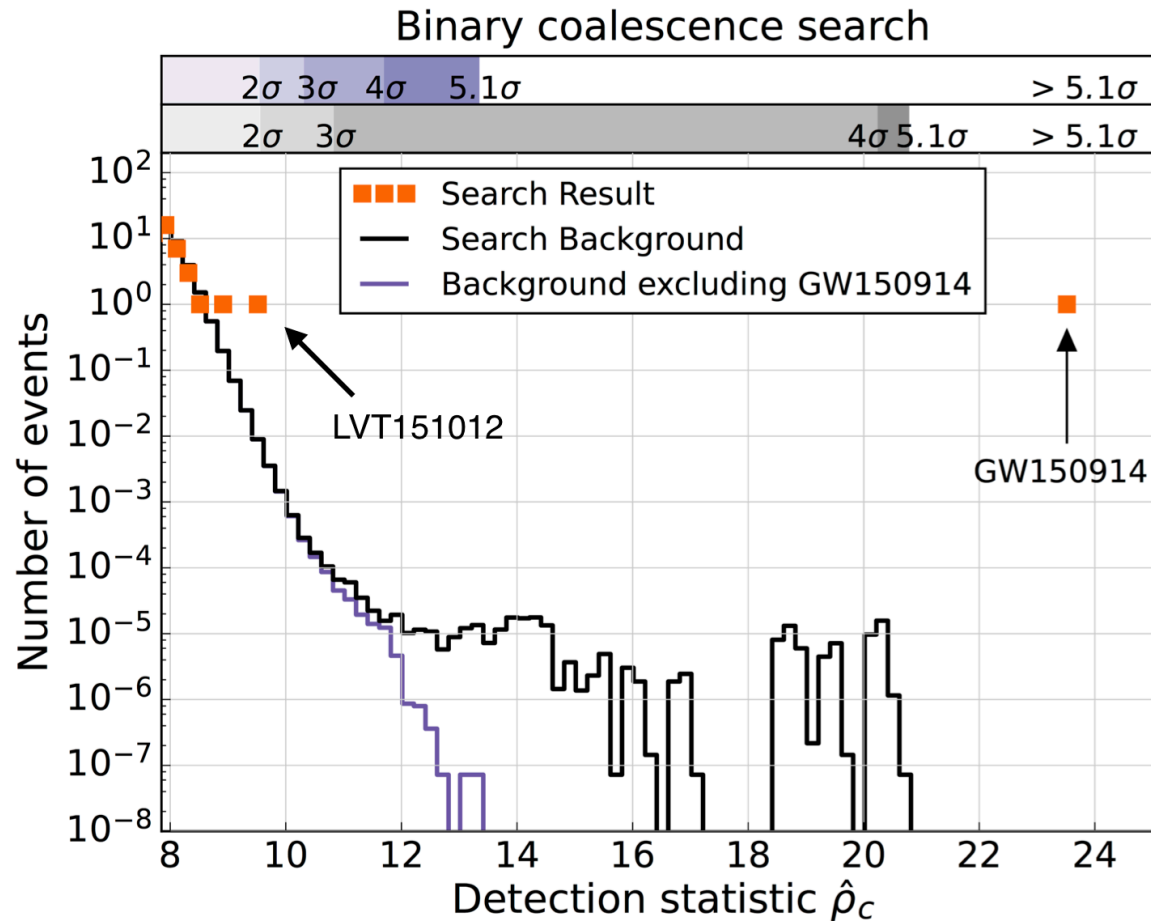
where d is the data from the detectors and h is the signal template

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

Hanford, Washington (H1)



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