Searches with PyCBC

Gareth Cabourn Davies

Gravitational Wave Open Data Workshop #4

May, 2021

DCC: LIGO Document G2101002-v1

Overview

- Introduction
- What are we looking for?
- What does the data look like?
- How do we search?
    - Matched filtering
    - Coincidence
    - Significance
- What if we are wrong in our signal assumptions?

Me, in case my camera doesn’t work

The Gravitational-Wave Spectrum

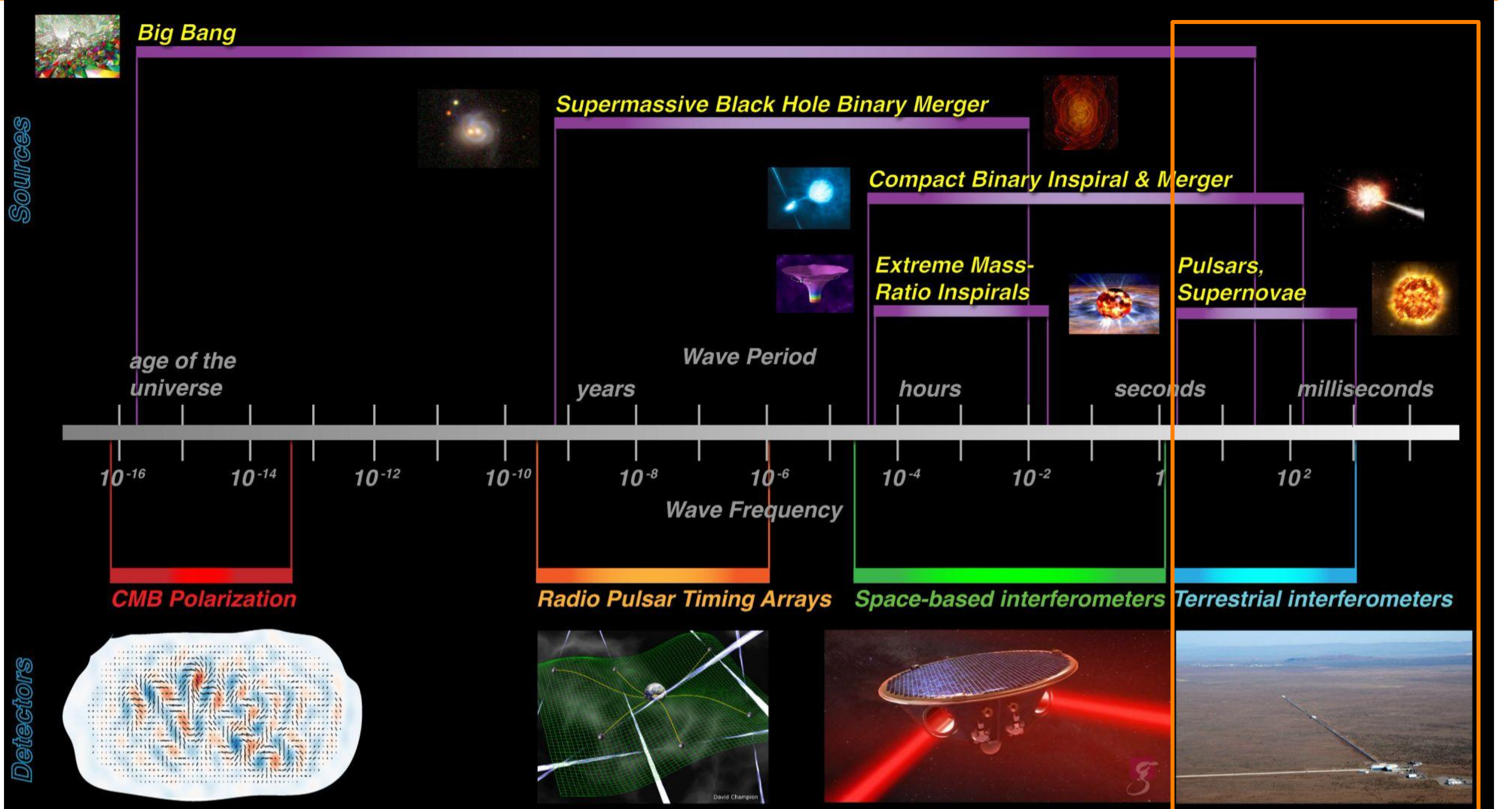


Image credit: Karan Jani/Georgia Tech

What we focus on

Compact Binary Coalescence

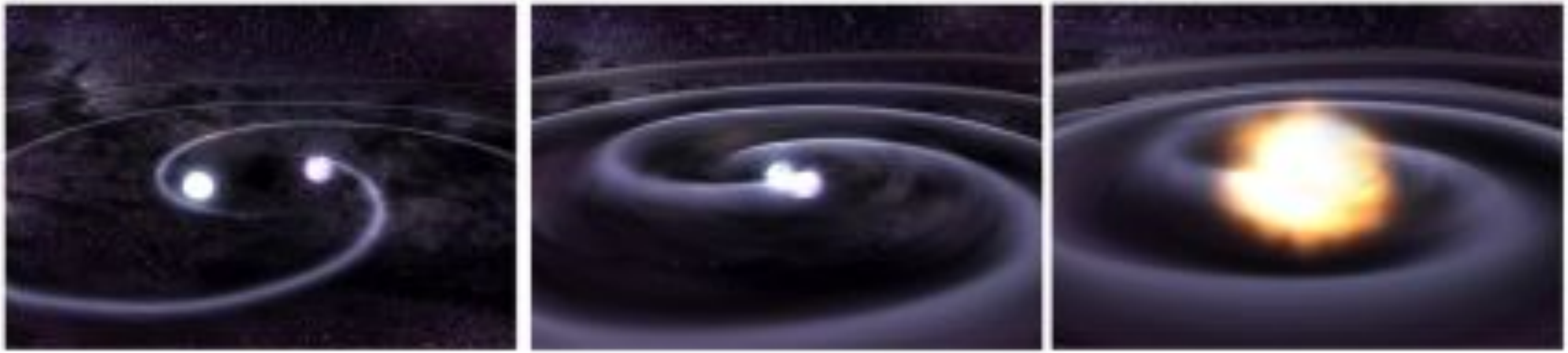


Image credit: NASA via ligo.org

Gravitational-wave event searches

There are two types of searches, online and offline

- Online searches are low-latency searches which aim to get quick results in order to get rapid alerts of events
- Offline searches use archived data using more computationally expensive techniques to get deeper searches into the data

What searches are there?

- Templated searches:
  - GstLAL - Online and Offline, lscsoft.docs.ligo.org/gstlal
  - **PyCBC** - Online and **Offline**, pycbc.org
  - MBTA - Online and Offline, T. Adams et al (2016)
  - SPIIR - Online only, Q. Chu (2017)
  - IAS - Offline only, Venumadhav et al. (2020)
- Non-templated search
  - cWB - Online and Offline gwburst.gitlab.io

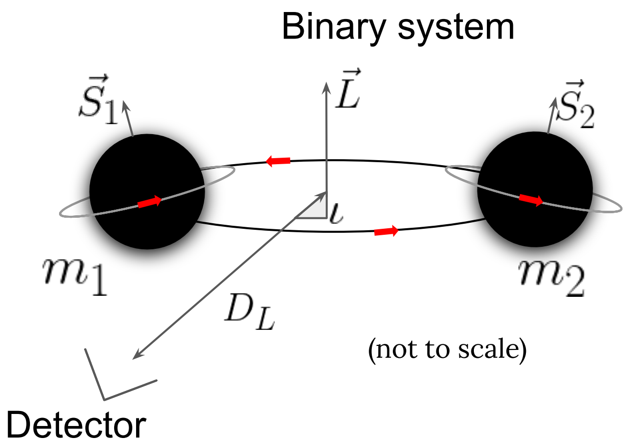
What are we looking for?

Modelling colliding black holes

What will the signals from these systems look like in the data?

The signal from a binary system made up of black holes will be described by fifteen parameters

- Intrinsic parameters:
  - Component Masses: m_1 m_2
  - Component spins in each direction: s_{1x} s_{1y} s_{1z} s_{2x} s_{2y} s_{2z}
- Extrinsic Parameters:
  - Location: Right Ascension and Declination
  - Inclination angle between line of sight and orbital plane, ι
  - Polarisation angle,
  - Phase at coalescence
  - Luminosity distance, D_L
  - Time of coalescence



Modelling colliding black holes

The above parameters and Einstein's GR equations exactly describe the dynamics of the system*

However these cannot be solved analytically - so we need to use approximate analytical solutions or numerical relativity

Analytical Solutions	Numerical solutions
• Perturbative approach can be used • Example: effective one body • Loses accuracy as closer to merger	• Directly solves equations • Very expensive • Can model collision • Some inaccuracy

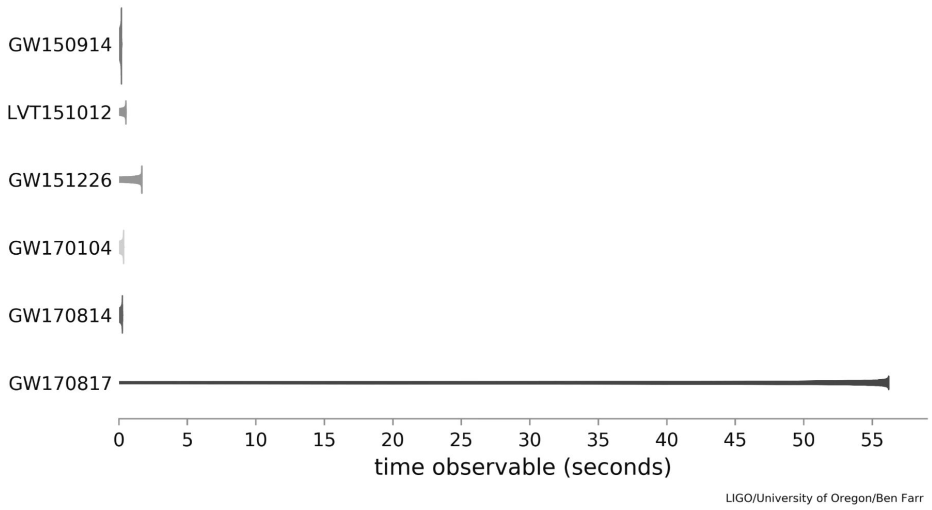
*Provided GR stands, e.g. Abbott et al Phys.Rev.D 100 (2019) 10, 104036

Waveform simplifications

In order to search the parameter space efficiently, we make certain assumptions about the systems to simplify the analysis

- The component masses and spins have most impact on waveform (m1, m2)
- We use templates with spins aligned with the binary angular momentum (s1z, s2z)
- We use face-on-binaries
- Location, polarisation, phase, distance and time can be reconstructed after the event is found
- Parameter Estimation can be used to reconstruct the waveform more accurately

Effect of masses on waveforms



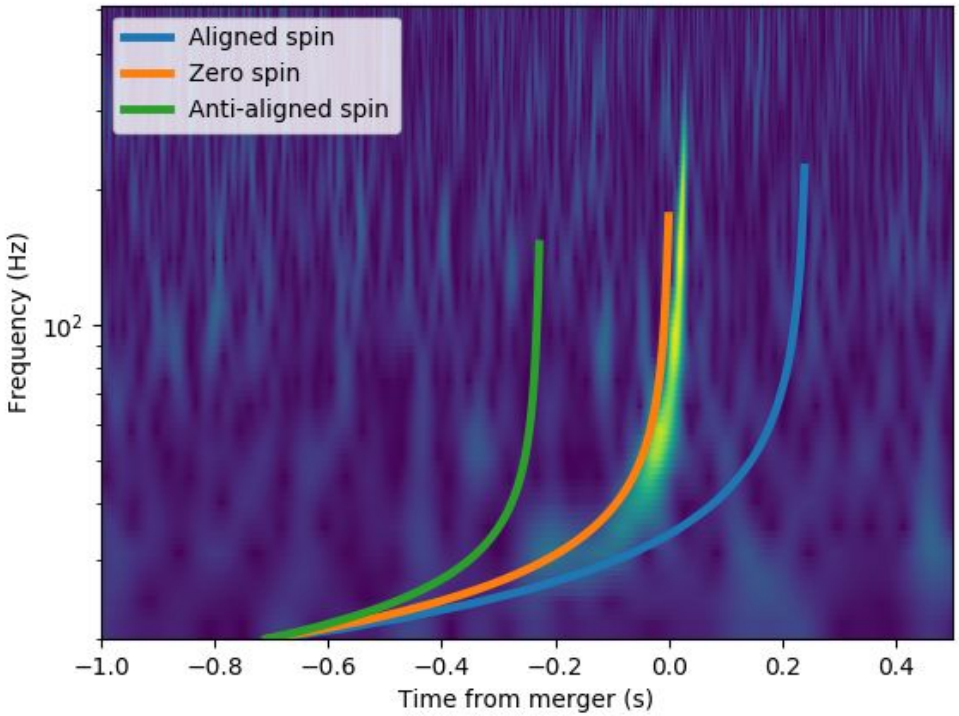
LIGO/University of Oregon/Ben Farr

Waveform simplifications

In order to search the parameter space efficiently, we make certain assumptions about the systems to simplify the analysis

- The component masses and spins have most impact on waveform (m1, m2)
- We use templates with spins aligned with the binary angular momentum (s1z, s2z)
- We use face-on-binaries
- Location, polarisation, phase, distance and time can be reconstructed after the event is found
- Parameter Estimation can be used to reconstruct the waveform more accurately

Effect of spins on waveforms



Signal model summary

- Developing and improving compact binary signal modelling is a large field of research, which has made very rapid progress
- Current waveform models are good enough for most purposes
- There are still areas for improvement (e.g. high-mass ratio signals, misaligned spins, extremal spins, exotic objects or non-GR waveforms)

What does the data look like?

Complicated noise curve

Many lines in the data, not such an issue for transient searches, but can be an issue for continuous wave searches

To an okay approximation, the detector data is colored Gaussian noise - standard Gaussian noise just with certain frequencies louder than others

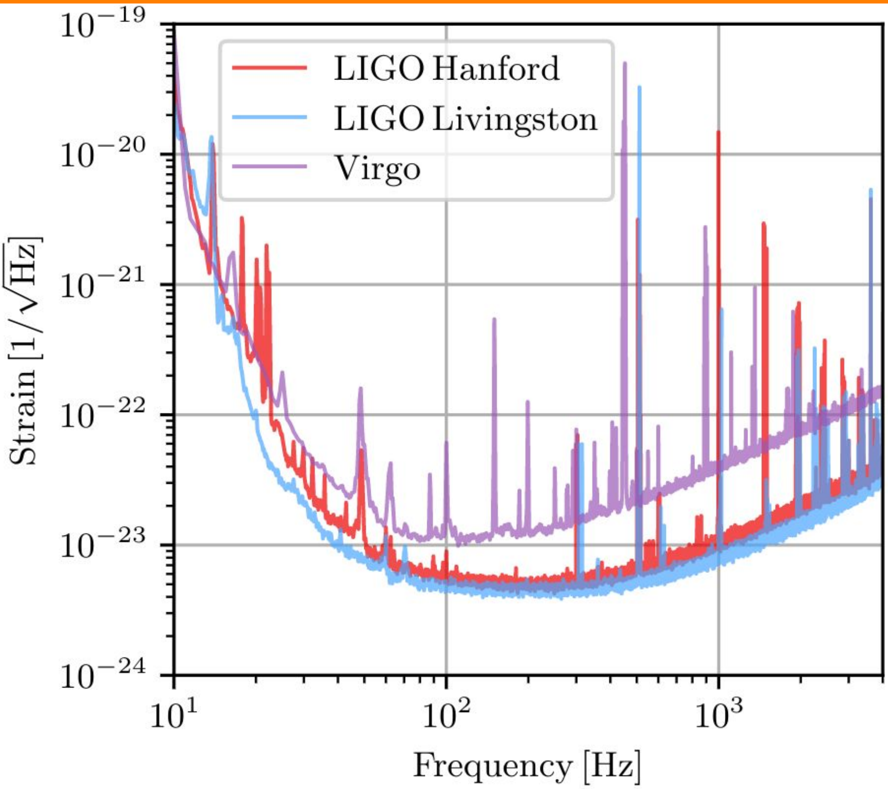


Image from Abbott et al (2020) GWTC-2 2010.14527

Non-stationarity

The detector sensitivity is not constant, this can happen rapidly or slowly

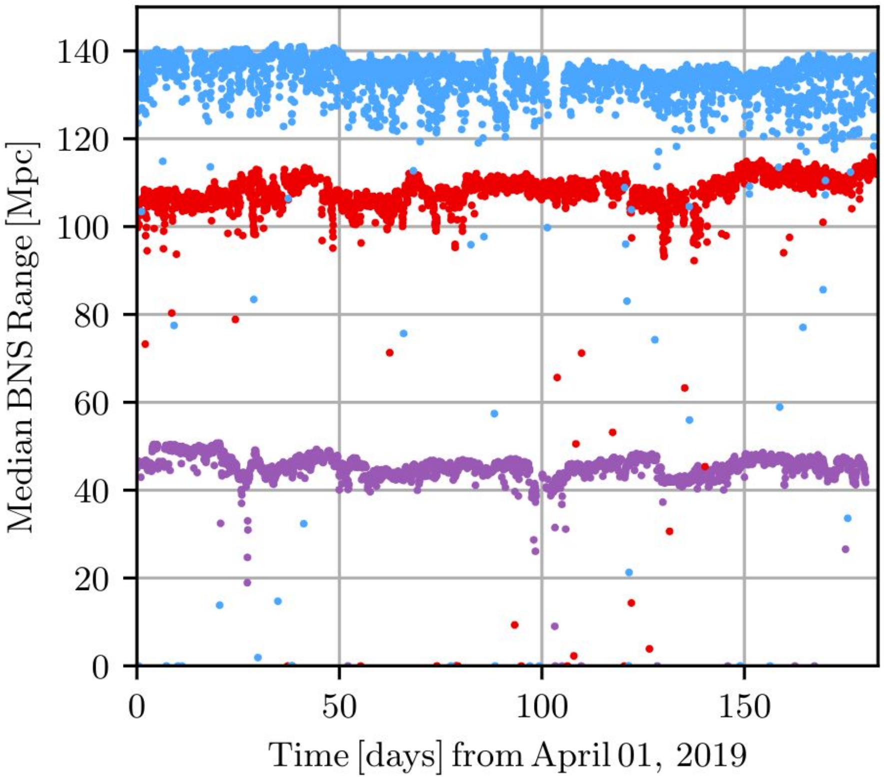
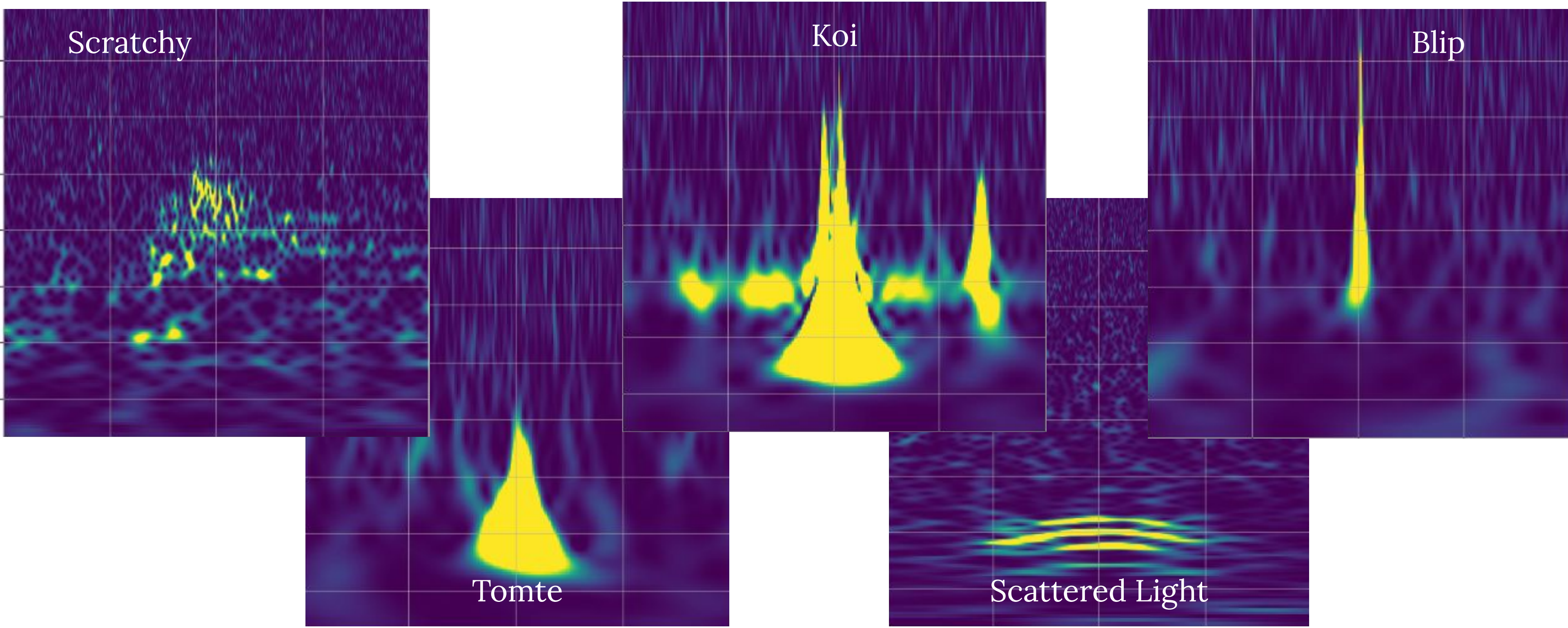


Image from Abbott et al (2020) GWTC-2 2010.14527

Non-Gaussian glitches



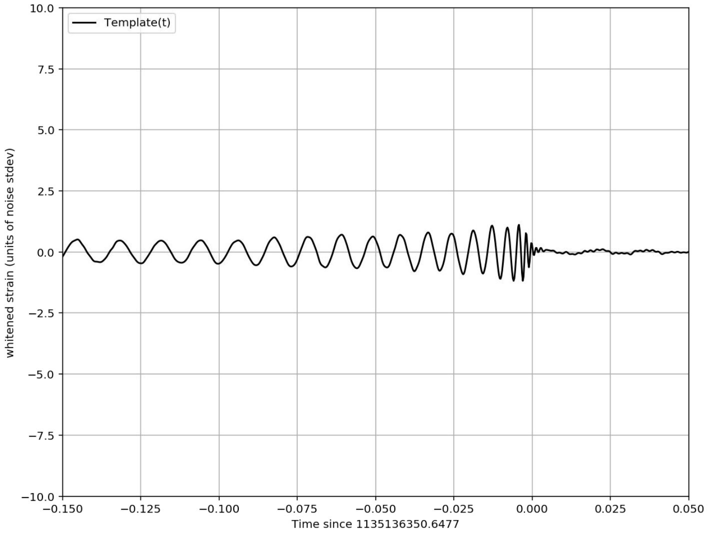
Detector noise summary

1. The noise curves are complex, with many lines
2. Sensitivity is highly non-stationary
3. Non-Gaussian artefacts regularly appear in the data

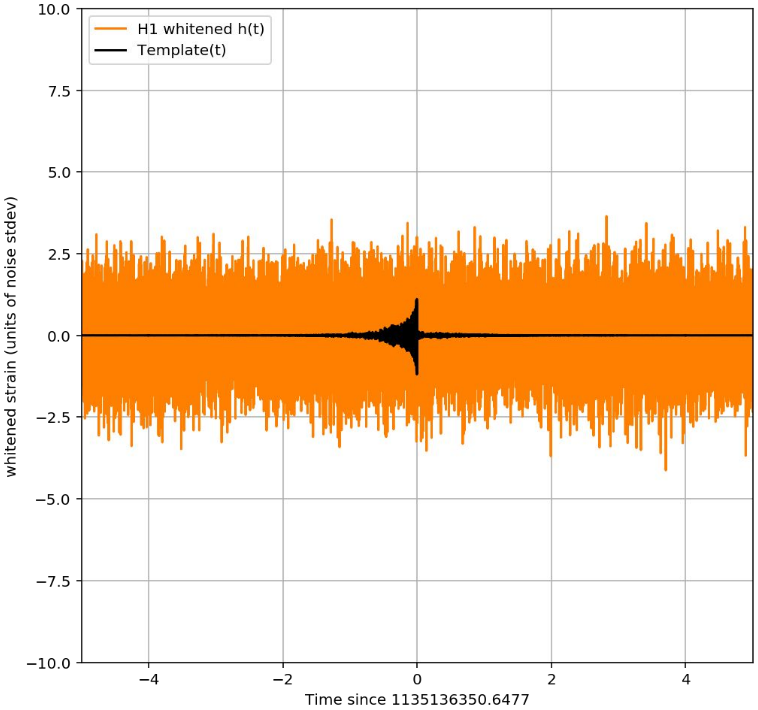
How do we search in the data?

Detection problem

We know what the signal looks like



But it is buried in detector noise



Adapted from GWOSC tutorial

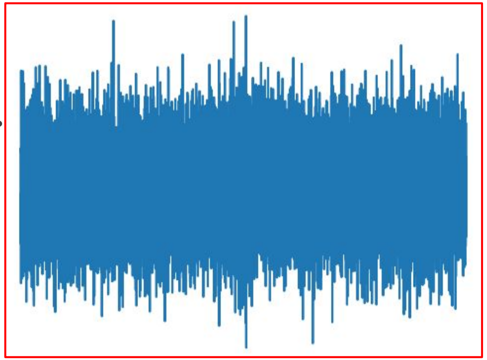
Matched Filtering

Optimal for signals:

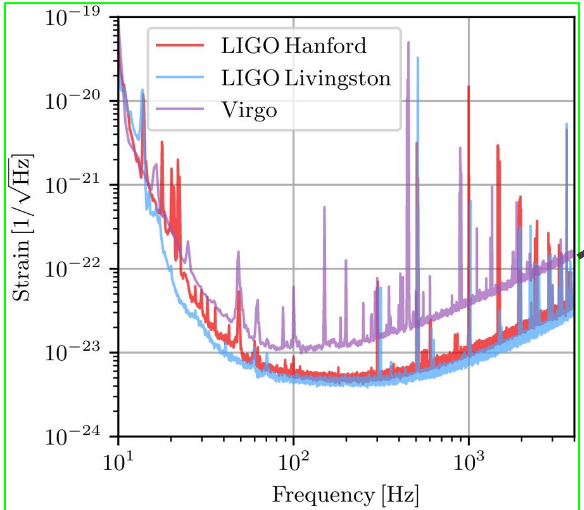
- in stationary Gaussian noise
- with known PSD

(Wainstein and Zubakov, 1962)

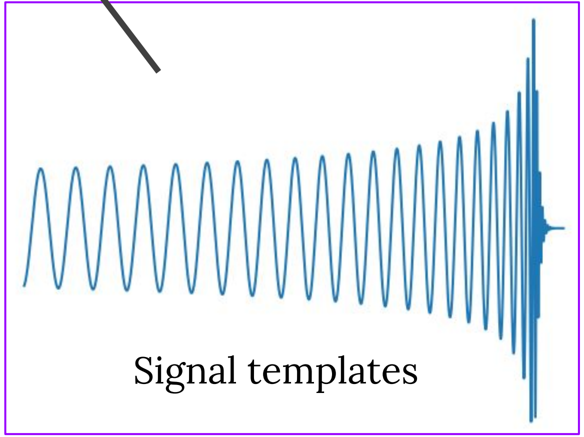
$$(s|h) = 4\Re \int_0^\infty \frac{\boxed{\tilde{s}(f)}\,\boxed{\tilde{h}^\star(f)}}{\boxed{S_h(f)}}\,\mathrm{d}f$$



The data



PSD



Signal templates

Modifications to SNR calculation

We want to maximise over some parameters and include others in our standard SNR calculation

$$(s|h) = 4\Re \int_0^\infty \frac{s(\tilde{f})\tilde{h}^\star(f))}{S_h(f)} \mathrm{d}f$$

Maximise over orientation and sky location

$$(s|h) = 4 \left| \int_0^\infty \frac{\tilde{s}(f)\tilde{h}^\star(f)}{S_h(f)} \mathrm{d}f \right|$$

Include coalescence time

$$(s|h) = 4 \left| \int_0^\infty \frac{\tilde{s}(f)\tilde{h}^\star(f)}{S_h(f)} e^{-2i\pi f t_c} \mathrm{d}f \right|$$

Which waveforms do we use?

As mentioned earlier, the parameters with most impact on the signal waveform are the masses and aligned spins of the components

We place templates within the bank randomly, but only if the match $(h|h)$ between templates is below a specific threshold.

This means that we end up with a bank which should match well to any signal within this parameter space

The template on the right has been used for the PyCBC-Broad search for many recent publications, and contains ~400k templates

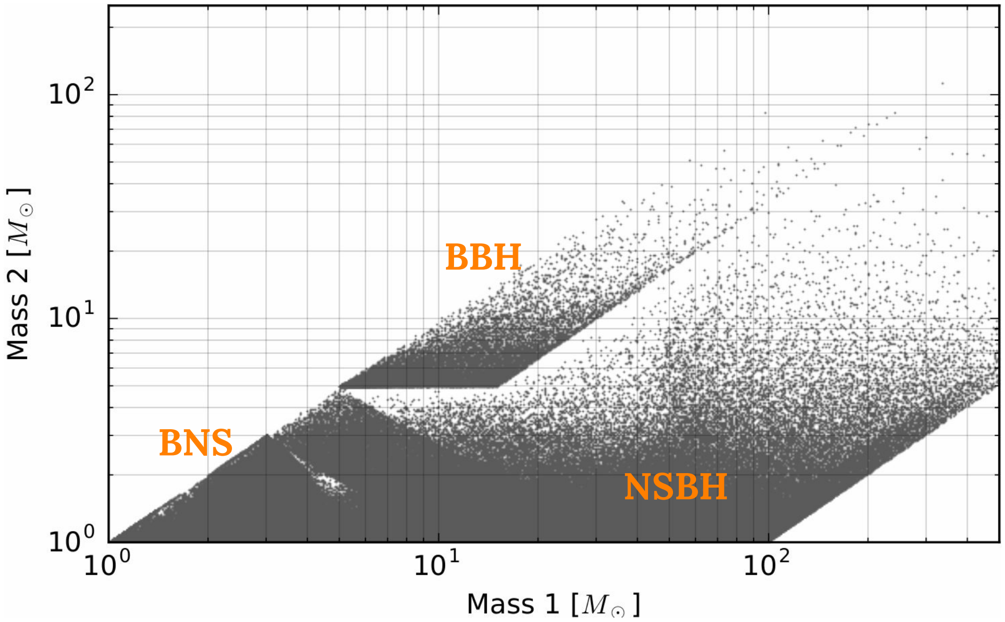
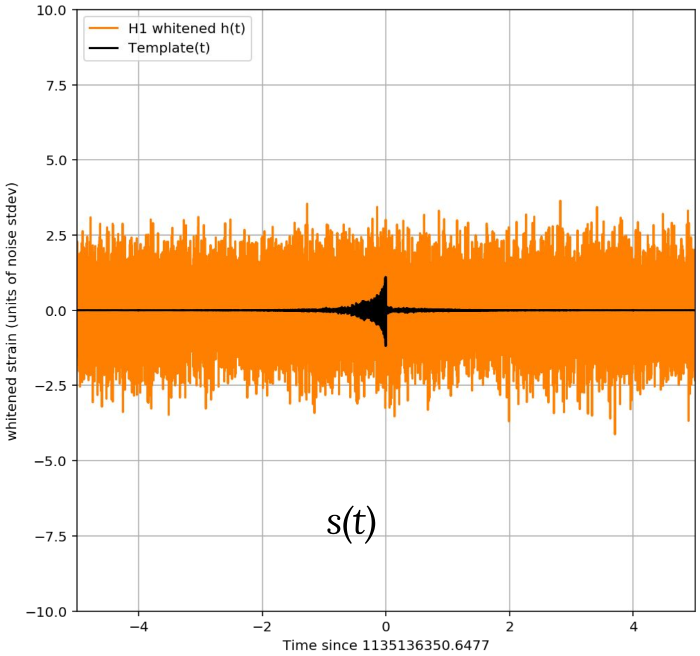
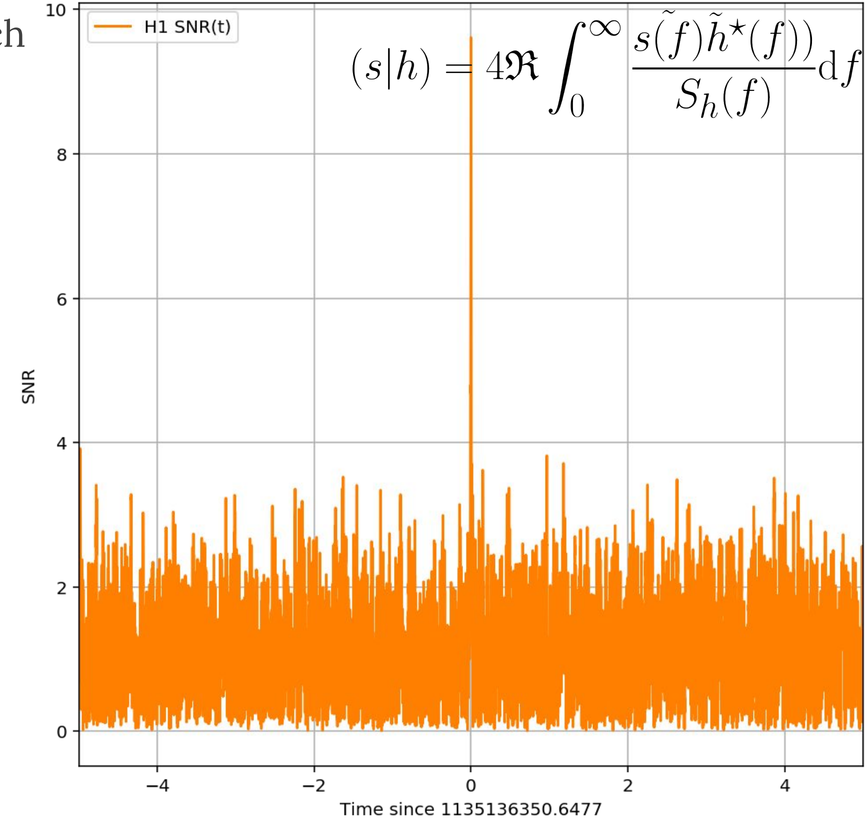


Image credit: Dal Canton and Harry (2017)

SNR time series

We end up obtaining a time series of SNR values for each template. The peaks in this time series are triggers



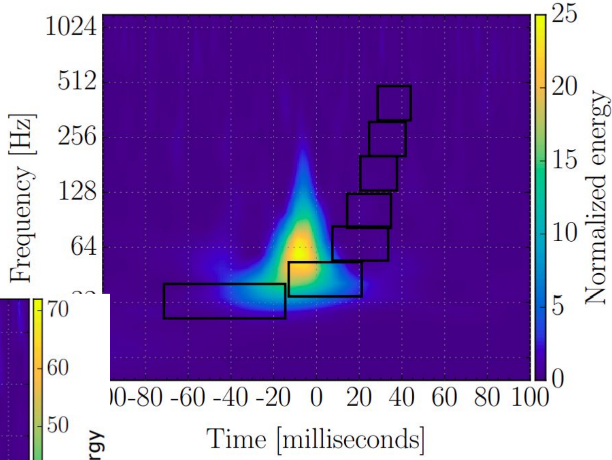


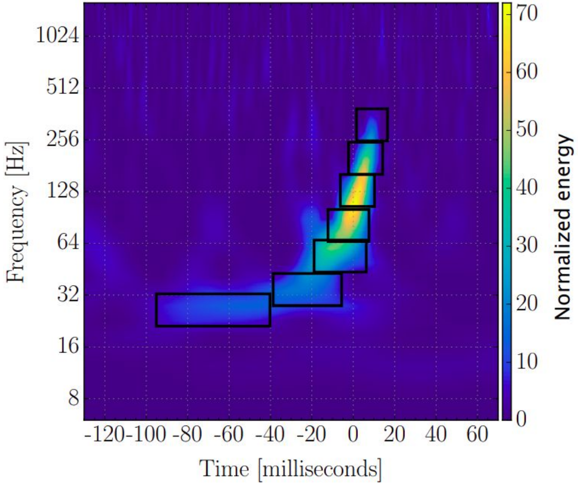
$$(s|h) = 4\Re \int_0^\infty \frac{s(\tilde{f})\tilde{h}^\star(f))}{S_h(f)} \mathrm{d}f$$

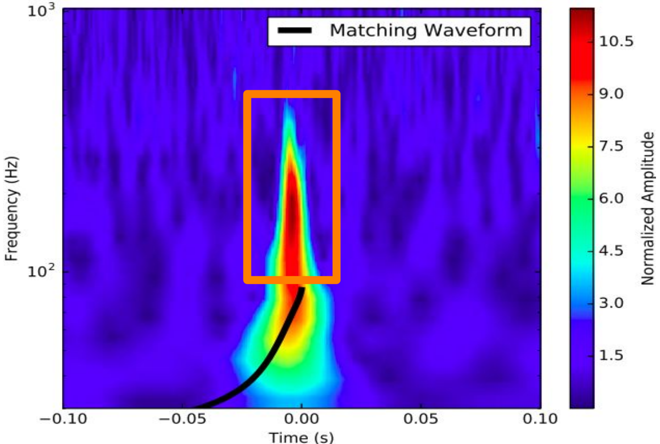
Beyond Matched Filtering

SNR is optimal *if data is Gaussian*. Data is not Gaussian

1. Split into frequency bins and check that the relative amount of power in each bin is correct (right)
2. Check for power above the final frequency of the signal (below)







Images: Nitz 2018 (left), Harry 2019 (above, ODW 1)

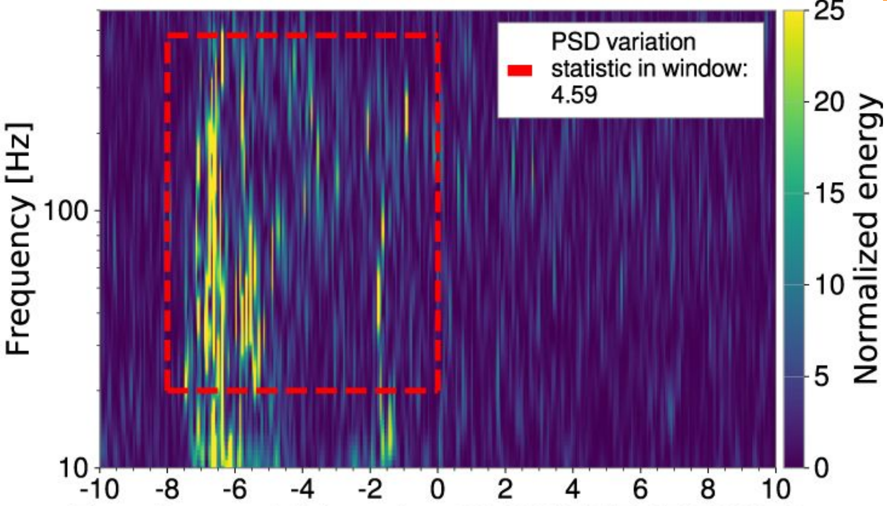
Beyond Matched Filtering

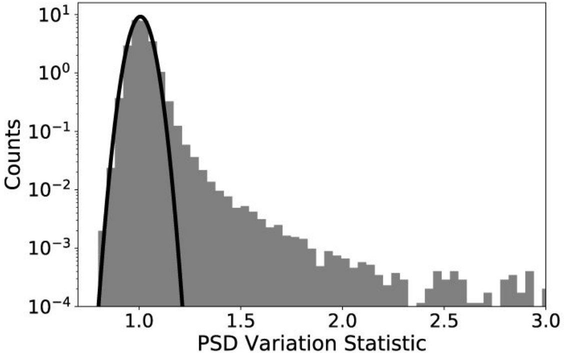
Basic idea to cope with non-stationarity is to keep remeasuring the power-spectral density (~every 512s)

Detectors can rapidly change sensitivity - this means that the PSD estimates used in matched filtering can be incorrect

Develop a test for how rapidly the PSD is changing, if it is changing too much, down-weight the trigger (or remove if really high)

S Mozzon et al (2020)

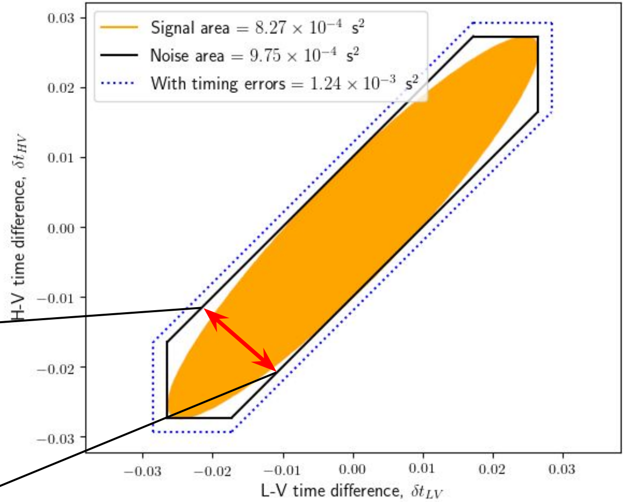


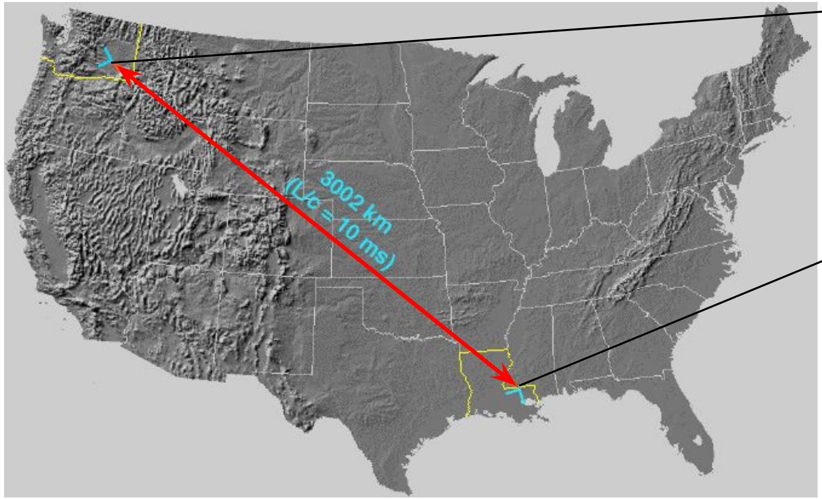


Coincidence

Noise triggers are not correlated between detectors

Therefore the fact that triggers are seen in multiple detectors simultaneously is a good discriminator





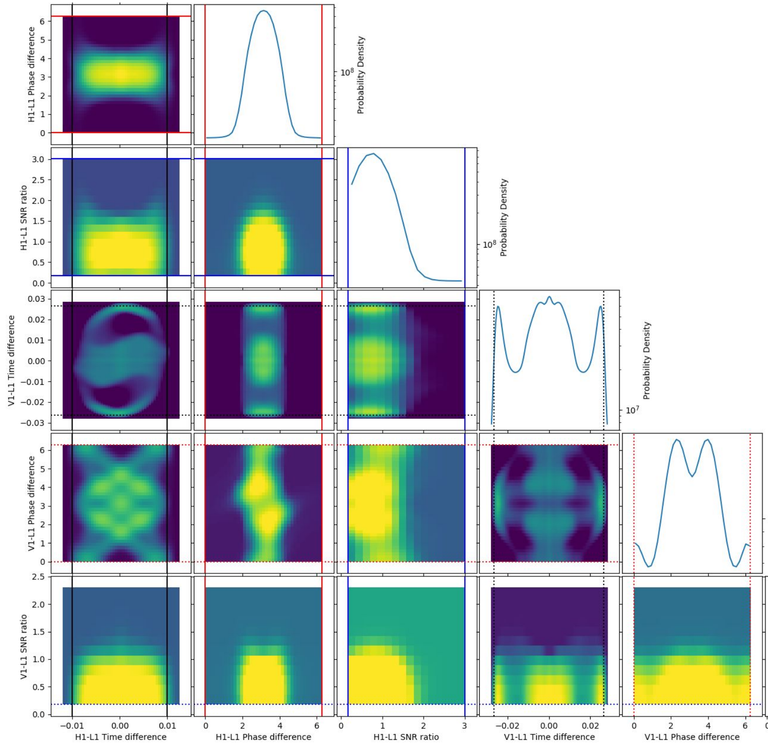
We insist that the triggers are within light-travel-time between each pair of coincident detectors (plus a bit extra for timing noise)

Image credits: G Sanders (2003 left), Davies et al (2020 above)

Beyond Coincidence

We use a ranking statistic based on the ratio of signal vs noise rate densities. This means we can incorporate extra information, e.g. tests for signal-like properties of the events. We check for:

- Are the time differences, SNR ratio and phase differences between triggers consistent with signals?
- Is the instantaneous sensitive volume bigger or smaller than usual?



Calculating Significance

We time-shift the data of one detector relative to the others

Coincidences in time shifts are our background

Assumes that noise triggers are not correlated between detectors (safe)

How many background triggers are ranked higher than the foreground? This is our false alarm rate

Do we include triggers from foreground events in the background?

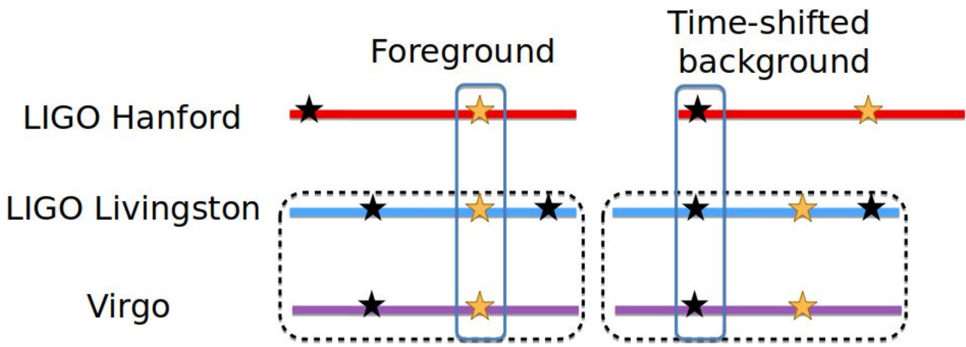


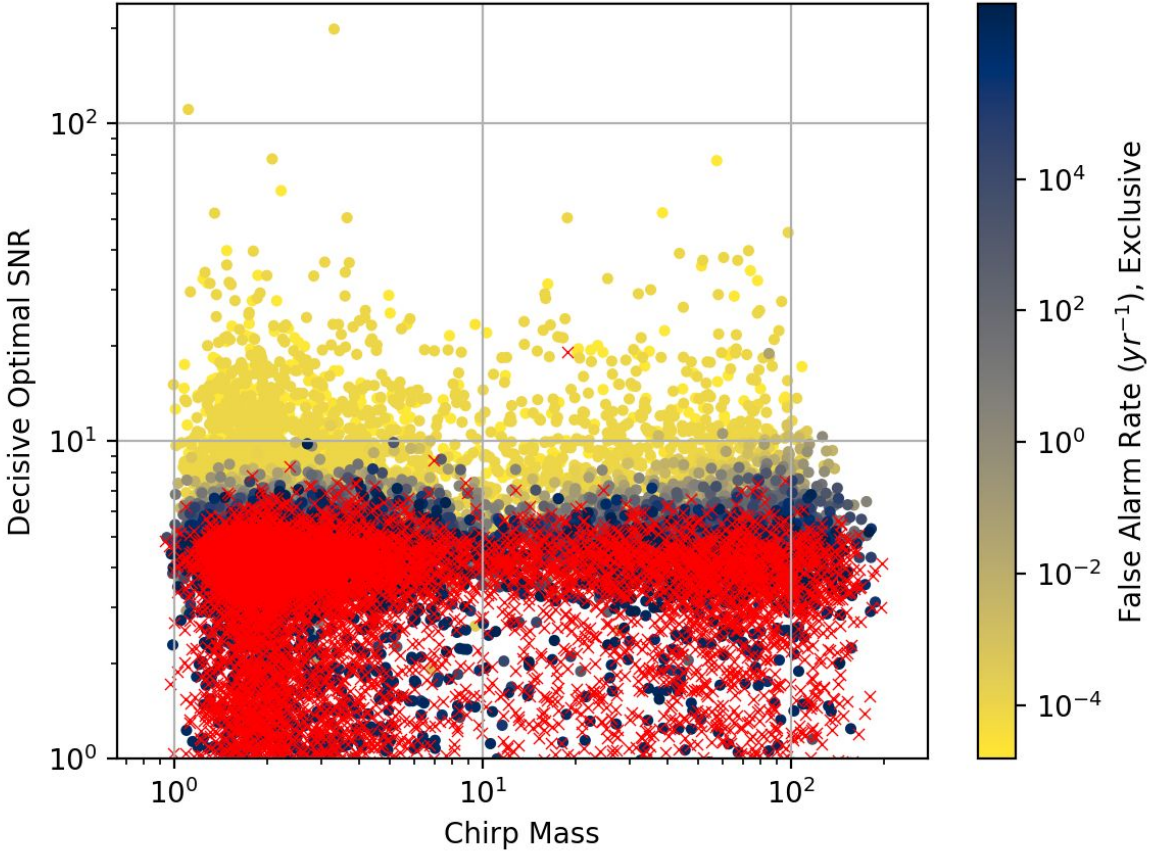
Image: Davies et al 2020

What if we are wrong?

Analysis validation

Lots of simulated signals

Manual checks for unusual data or background features before checking actual results



How much effort?

We search a lot of templates, and do a lot of background analysis - need to be parallel

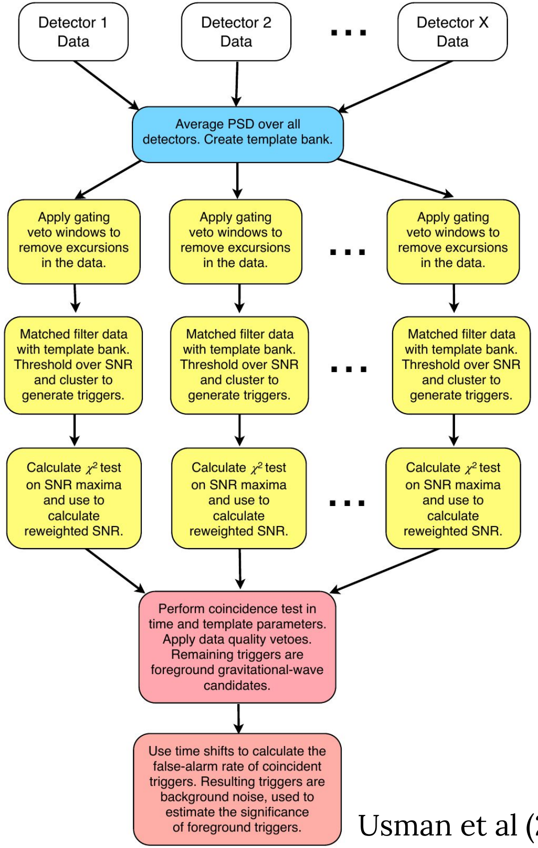
PyCBC searches can use:

- Open science grid: https://opensciencegrid.org
- GPUs

Condor workflows managed by Pegasus

HTCondor High Throughput Computing

Pegasus



Usman et al (2016)

Weakly-modelled searches

- We don’t only rely on matched-filtering
- Our search makes a number of assumptions
- Maybe our waveform models are wrong?
- Maybe general relativity is wrong?
- Maybe we have astrophysical sources that were not expected, or are not easily modelled (supernovae)?

Basic idea of burst searches

Example: cWB Klimenko et al. (2008)

- Create q-transform spectrograms of data at all times (Remember Laura Nuttall's talk for q-transform explanation)
- Look for features standing out from the noise
- Look for consistent morphology in both observatories
- We can impose CBC-like morphology, e.g. increasing frequency with time

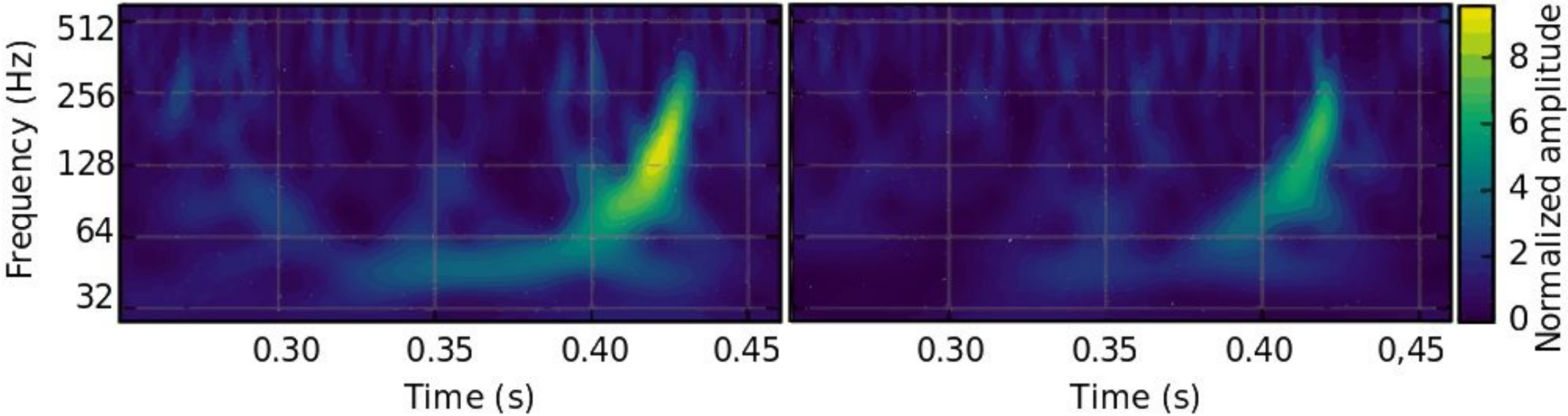
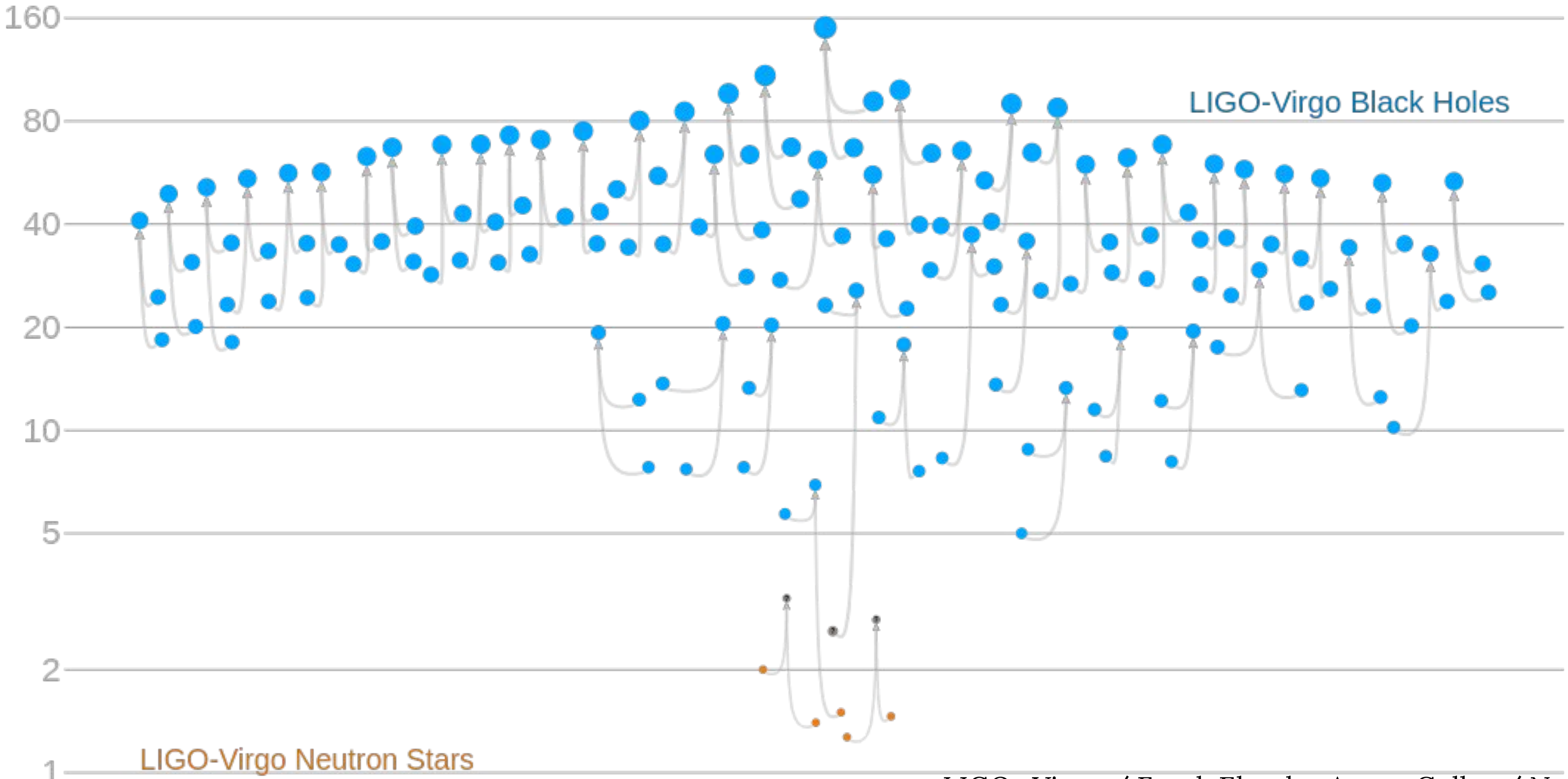


Image: B.P Abbott et al Phys. Rev. Lett. 116, 061102 (2016)

Look at all these black holes



LIGO -Virgo / Frank Elavsky, Aaron Geller / Northwestern

Conclusion

- Predictions from GR allow us to search for gravitational waves from compact binary mergers using large numbers of waveform templates
- LIGO-Virgo noise features present challenges for identification of gravitational-wave signals
- Current searches rely on matched-filtering, with signal tests to account for non-Gaussianities
- Also use unmodelled searches to catch the unexpected
- We have found lots already - let’s find more!

References

[1] T. Adams et al., (2016) Class. Quant. Grav., 33 175012
[2] F Aubin et. al. (2021) Class.Quant.Grav. 38 9, 095004
[3] Q. Chu, Ph.D. Thesis, The University of Western Australia. (2017)
[4] T Dal Canton and IW Harry (2017) e-Print: 1705.01845
[5] T Dal Canton et al (2020) e-Print: 2008.07494
[6] **GS Davies** et al (2020) Phys.Rev.D 102 2, 022004
[7] S Hooper et al (2012) Phys.Rev.D 86 024012
[8] Klimenko et al., (2008) CQG. 25 114029
[9] S Mozzon et al (2020) Class. Quantum Grav. 37 215014
[10] AH Nitz et. al. (2020) Astrophys.J. 891 123
[11] AH Nitz (2018) Class.Quant.Grav. 35
[12] T Venumadhav (2020) et. al. Phys.Rev.D 101 8, 083030
[13] S. A. Usman et al. (2016) Class. Quant. Grav. 33 215004
[14] Wainstein and Zubakov (1962) Extraction of signals from noise

Questions?