Data quality in gravitational-wave detectors

Ronaldas Macas
(ronaldas.macas@port.ac.uk)

16th May, 2023

Acknowledgements

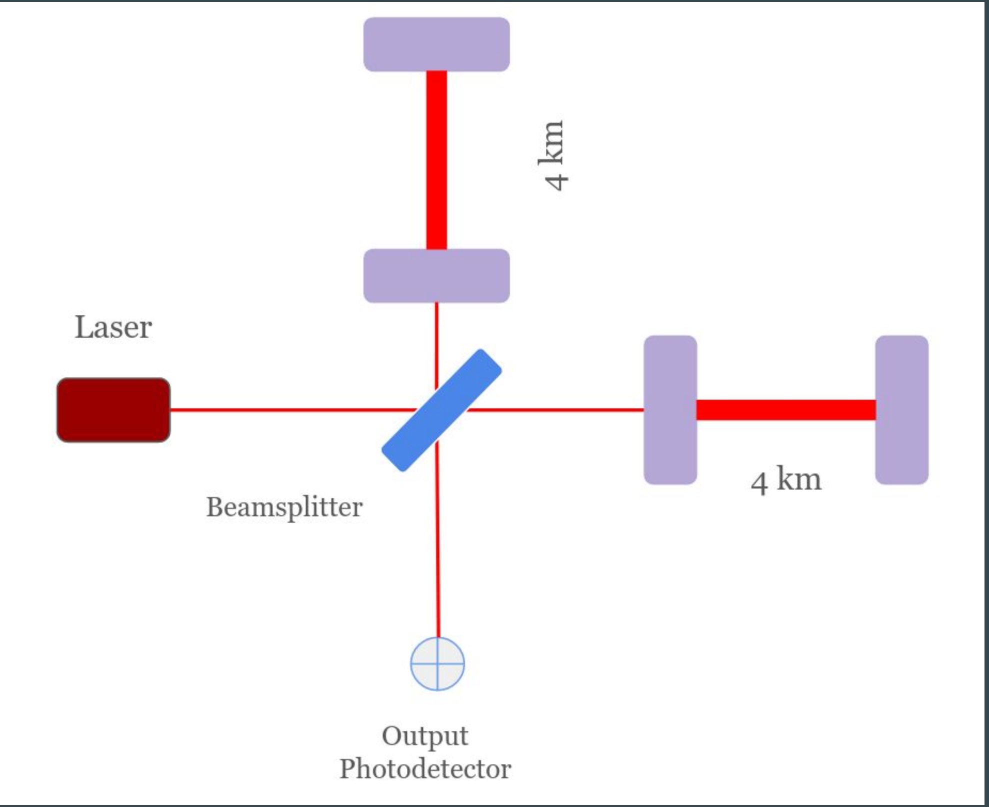
Many slides are adapted from data quality presentations made by Siddharth Soni, Marissa Walker, Duncan Macleod and Jess McIver.

For previous workshop slides, see https://gwosc.org/odw/

Outline

- Strain data
    - Time domain
    - Frequency domain
    - Time-frequency representation
- Data quality
    - Noise artifacts
    - Tools to inspect data
    - O3 and O4
- References (important!)

What is strain data *h(t)*?

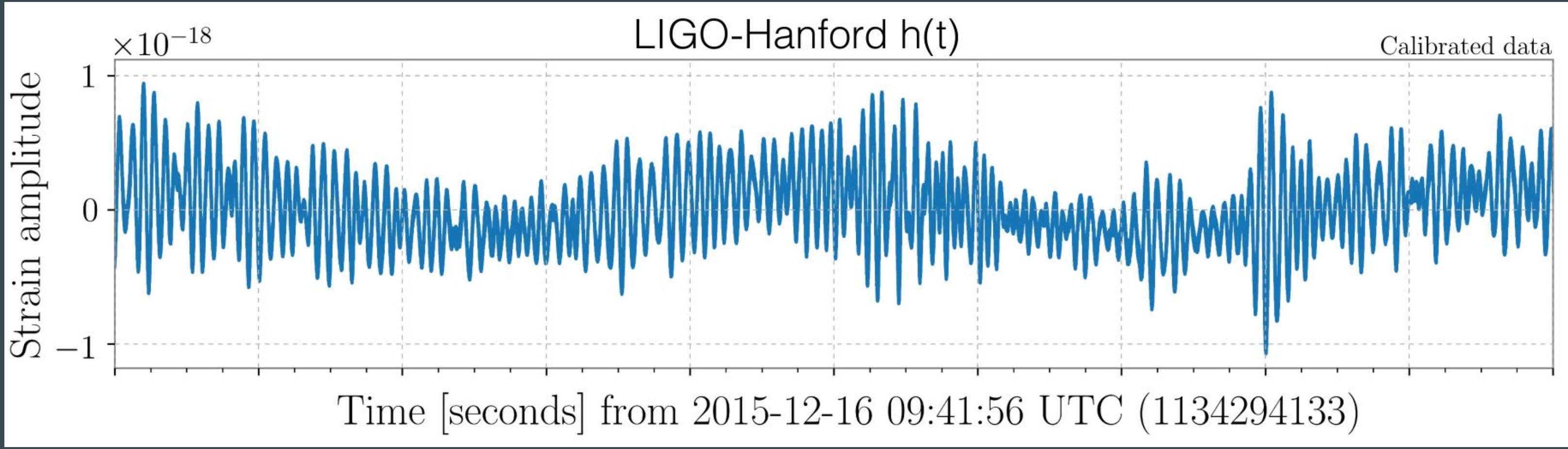


GW strain *h(t)* is the relative difference between L_x and L_y arms

$$h = \frac{L_x - L_y}{L}$$

Very simplified! More detailed detector layout will be discussed in slide 19.

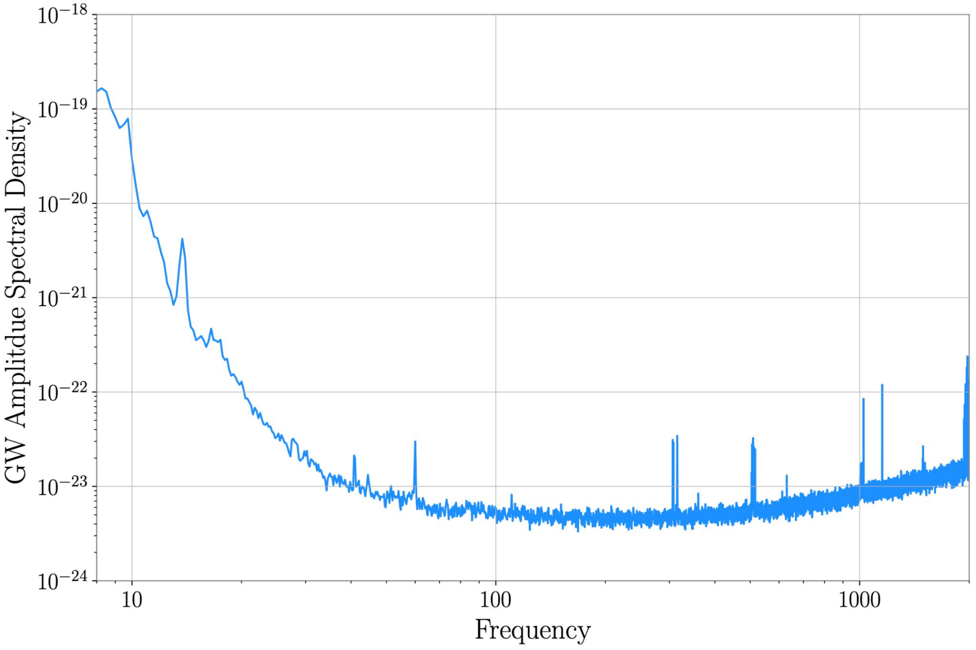
Raw time series data



- *h(t)* sampling rate for LIGO detectors: 16384 Hz
    - Open data: 16384 or 4096 Hz
- Looks really complicated!
    - We will with deal with that later...

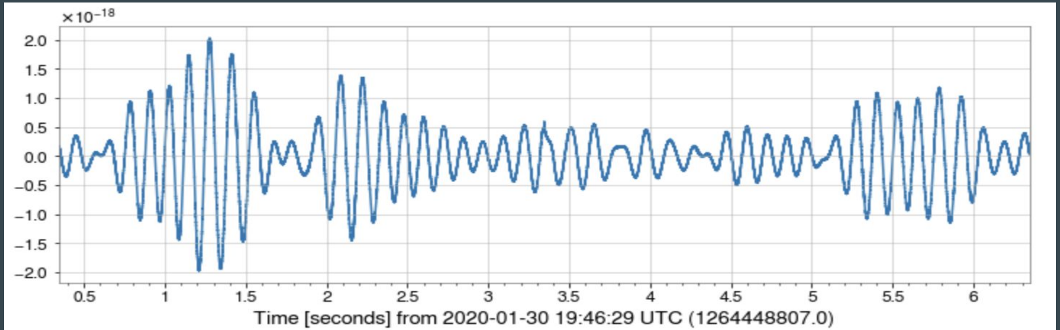
Frequency series data

1. Take time-series data (e.g. 512s)
2. Fourier-transform short segments of the time-series data (e.g. 4s)
3. Take the median Fourier transform
    - This is the median detector sensitivity

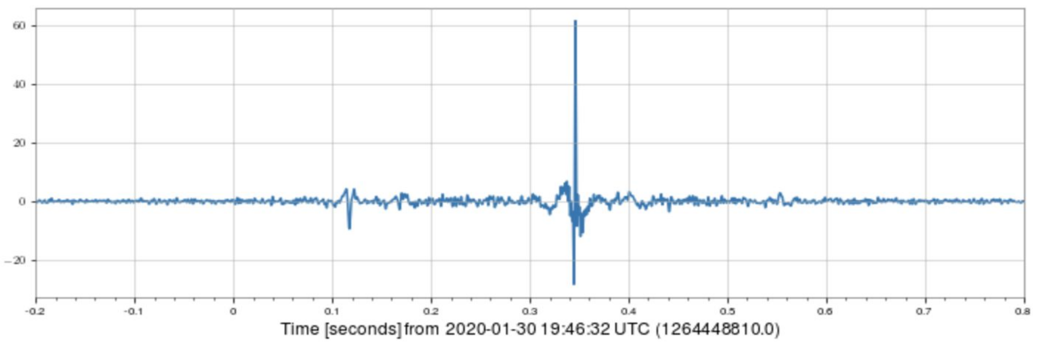


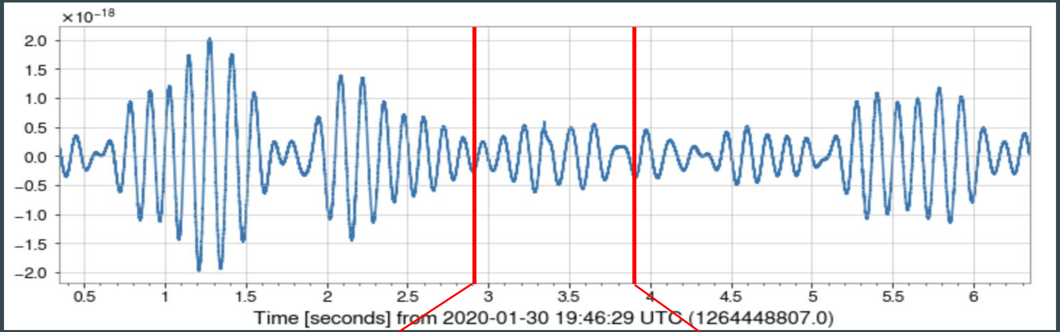
Whitening the time series

- Transforming strain data to frequency domain allows to estimate the average detector sensitivity for each frequency bin
    - This sensitivity is called amplitude spectral density (ASD)
- Having ASD allows us to “whiten” the data
    - In other words, “scale the data”
- For example: detector is less sensitive at lower frequencies (<20 Hz), so the data at low frequencies should be “less important” than at medium frequencies (20-100 Hz)

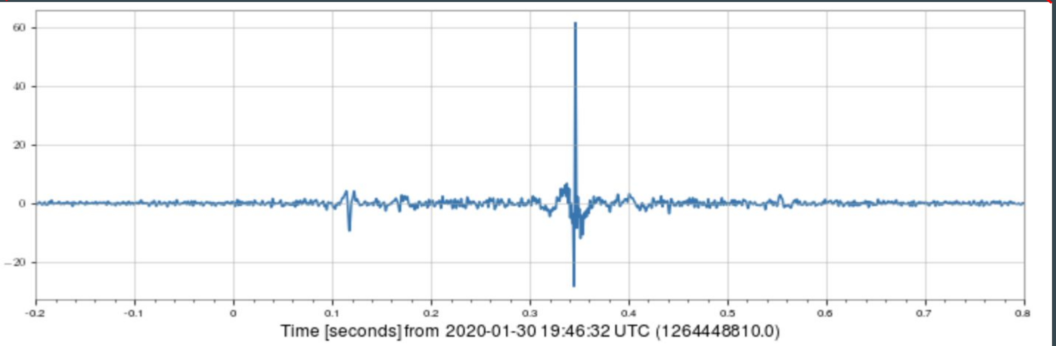


Divide by ASD



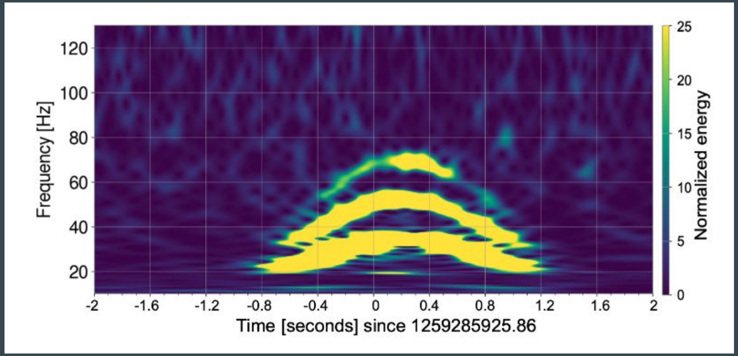


Divide by ASD

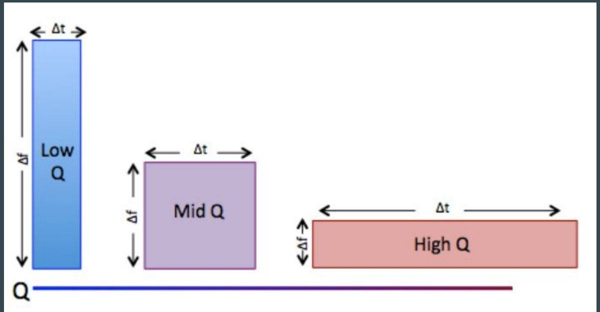


Time-frequency representation: Q-transform

- GW scientists often use time-frequency representation to inspect the data visually
  - Use Q-transform / Qscan / Qseries/ spectrogram/ omegascan, …
- Q-transform
  - Select Q-value
  - “Tile” the data for various Q values
  - Find the most optimal Q value
  - Make a Q-transform plot for this Q value



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

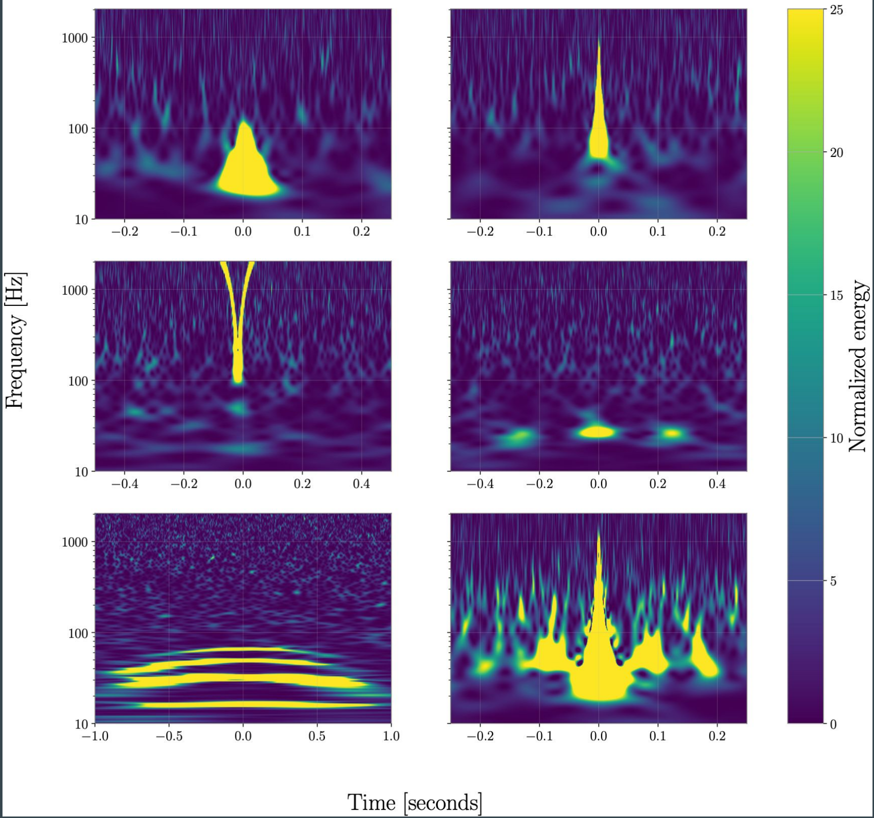


Outline

- Strain data
    - Time domain
    - Frequency domain
    - Time-frequency representation
- Data quality
    - Noise artifacts
    - Tools to inspect data
    - O3 and O4
- References (important!)

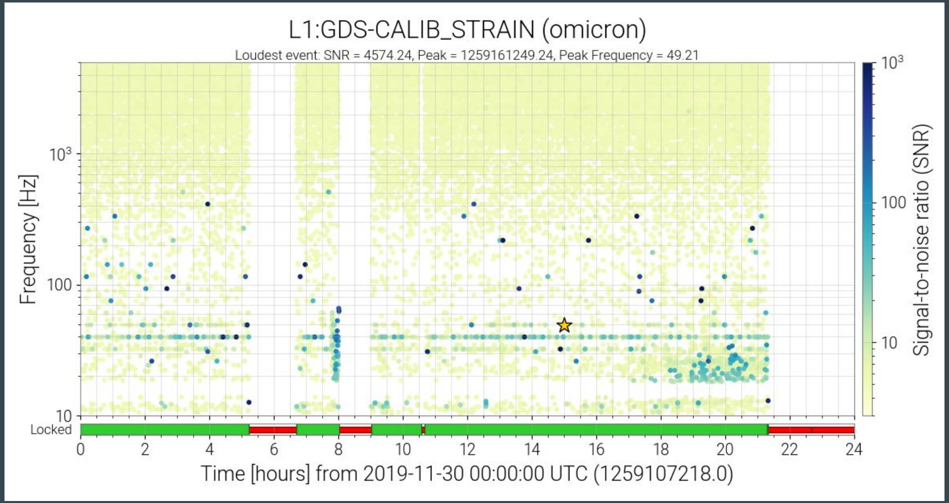
Gravitational-wave noise

- Gravitational-wave (GW) data is non-Gaussian and non-stationary
  - It contains noise artifacts ("glitches")
- Glitches can affect
  - GW detector sensitivity
  - GW searches
  - source parameter estimation, e.g. sky localisation



Gravitational-wave noise

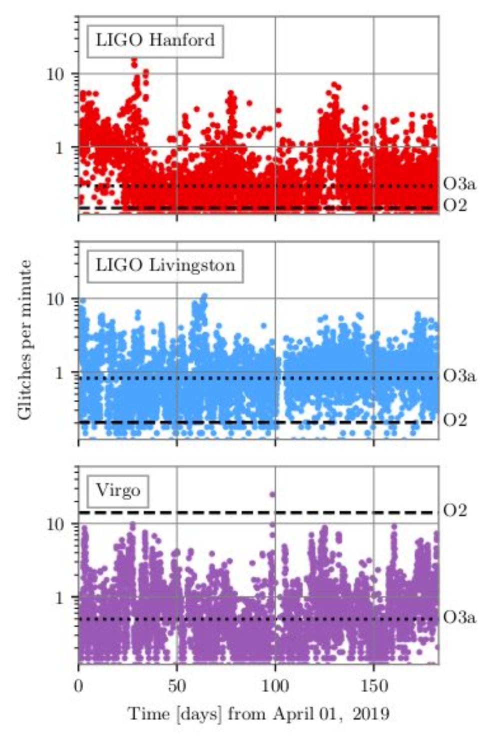
- Gravitational-wave (GW) data is non-Gaussian and non-stationary
    - It contains noise artifacts ("glitches")
- Glitches can affect
    - GW detector sensitivity
    - GW searches
    - source parameter estimation, e.g. sky localisation
- Glitches are detected using Omicron[1]



[1] Robinet et al. (2021)

Gravitational-wave noise

- Gravitational-wave (GW) data is non-Gaussian and non-stationary
  - It contains noise artifacts ("glitches")
- Glitches can affect
  - GW detector sensitivity
  - GW searches
  - source parameter estimation, e.g. sky localisation
- Glitches are detected using Omicron[1]
- Glitch rate varies over time

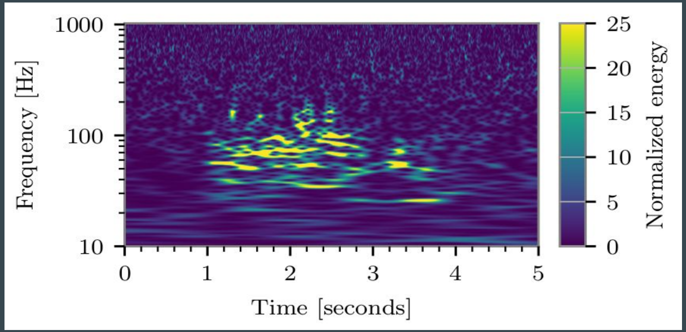


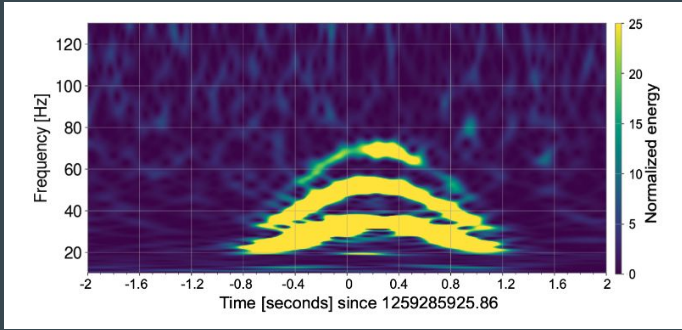
LIGO-Virgo (2021)

[1] Robinet et al. (2021)

Origin of glitches

- Origin of some glitches are known
  - Natural, e.g. thunderstorms
  - Human-made, e.g. trains or a fridge connected to the main power (aLOG: 23483)
- Some of glitches are recorded by witness channels
  - e.g. light scattering but not blips
- Knowing the origin of glitches allows to remove or mitigate them
  - Implemented RC tracking to reduce light scattering

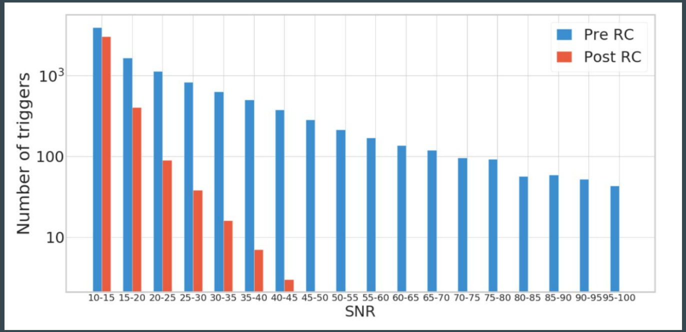


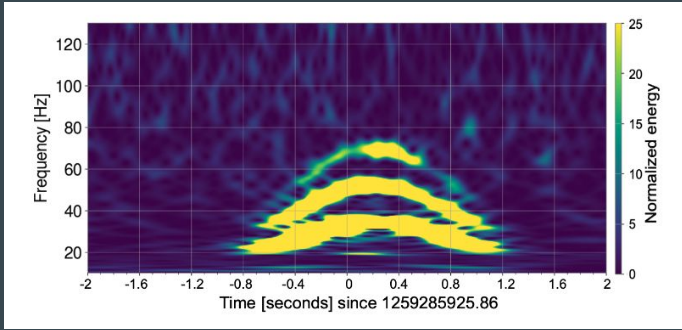


Soni et al. (2021)

Origin of glitches

- Origin of some glitches are known
  - Natural, e.g. thunderstorms
  - Human-made, e.g. trains or a fridge connected to the main power (aLOG: 23483)
- Some of glitches are recorded by witness channels
  - e.g. light scattering but not blips
- Knowing the origin of glitches allows to remove or mitigate them
  - Implemented RC tracking to reduce light scattering





Soni et al. (2021)

Data inspection tools

- We want to get rid of glitches but how?
  - Identify the noise
  - Look for potential correlations with the witness channels
  - Perform tests to simulate the noise
  - Fix the source of noise to reduce or eliminate it
  - If this cannot be done, try modelling the noise or create vetoes

- Data inspection tools used by the LVK
  - Omicron
  - Q-transform
  - GravitySpy
  - Hveto
  - Detector status pages

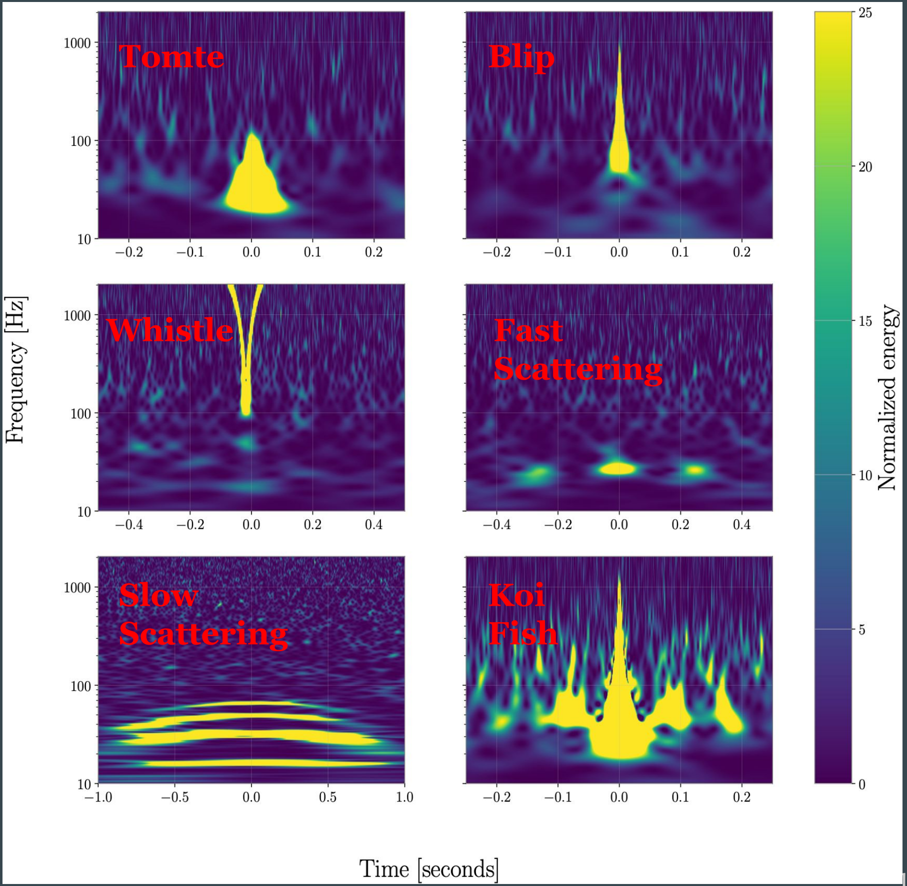
Data inspection tools

- We want to get rid of glitches but how?
  - Identify the noise
  - Look for potential correlations with the witness channels
  - Perform tests to simulate the noise
  - Fix the source of noise to reduce or eliminate it
  - If this cannot be done, try modelling the noise or create vetoes

- Data inspection tools used by the LVK
  - Omicron
  - Q-transform
  - GravitySpy
  - Hveto
  - Detector status pages

Gravity Spy[2]

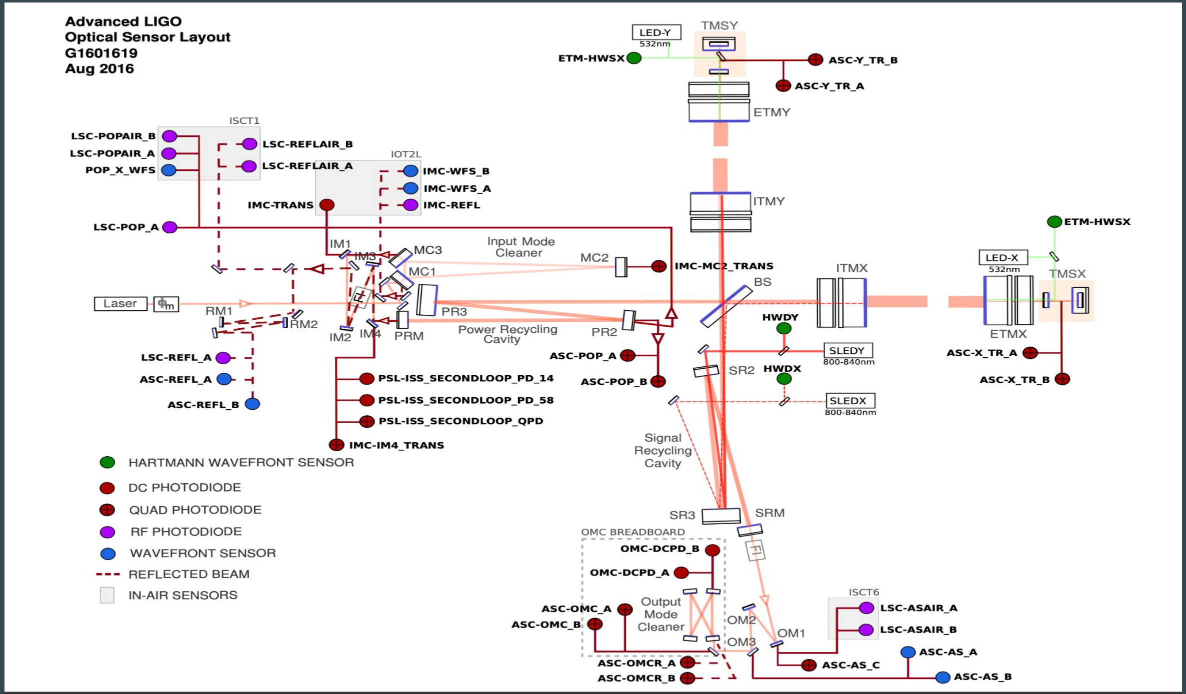
- An image recognition algorithm based on convolutional neural networks
- Classifies transient noise at LIGO in 23 classes
- The algorithm is trained on time-frequency spectrograms of noise transients
- Uses Omicron triggers as the input and the output is predicted glitch class



[2] Zevin et al. (2017)

Witness channels

- GW detectors have thousands of sensors that record various activity



Hierarchical veto (Hveto)[3]

- Statistical correlations between noise in GW strain channel and witness channels
- Allows to find the potential noise culprits
- Does not work all the time!
  - Some noise sources are not recorded by any witness channels...

Summary

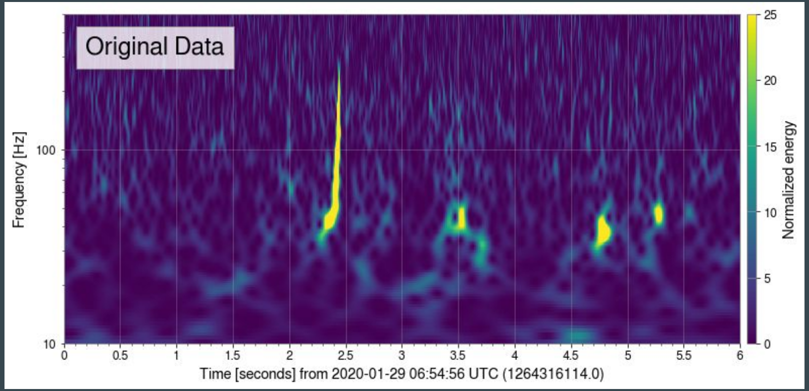
Summary of this HierarchichalVeto analysis.

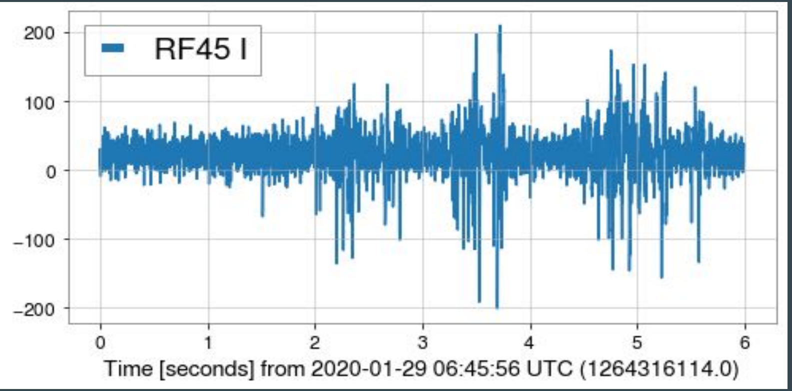
Round	Winner	Twin [s]	SNR Thresh	Significance	Use [%]	Efficiency [%]
1	L1:LSC-REFL_A_LF_OUT_DQ	0.10	10.00	42.82	13.43 [36/268]	1.20 [36/3011]
2	L1:SUS-ETMX_L3_OPLEV_SUM_OUT_DQ	1.00	15.00	10.39	8.17 [77/942]	2.59 [77/2975]
3	L1:ASC-X_TR_A_NSUM_OUT_DQ	1.00	9.00	8.54	32.50 [13/40]	0.45 [13/2898]
4	L1:PEM-CS_ACC_HVAC_FLOOR_Z_DQ	0.80	9.00	7.91	11.80 [44/373]	1.14 [33/2885]
5	L1:ASC-CHARD_P_OUT_DQ	0.10	12.00	7.70	12.77 [6/47]	0.21 [6/2852]
6	L1:LSC-POP_A_RF9_Q_ERR_DQ	0.80	7.75	6.26	6.25 [42/672]	1.51 [43/2846]
7	L1:ASC-Y_TR_B_PIT_OUT_DQ	1.00	7.75	5.84	5.86 [86/1468]	3.07 [86/2803]
8	L1:PEM-MY_ACC_BEAMTUBE_1900Y_Y_DQ	1.00	8.00	5.64	9.60 [29/302]	1.03 [28/2717]

[3] Smith et al. (2011)

Noise modelling

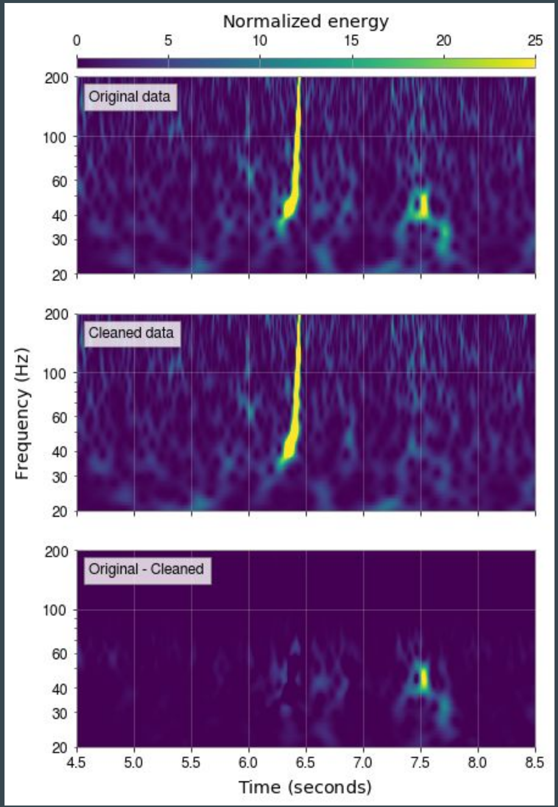
- Quite often the root cause of noise cannot be found
- Instead, try to model the noise





Noise modelling

- Quite often the root cause of noise cannot be found
- Instead, try to model the noise
    - Linear noise subtraction[4]
        - Use witness channel to model the noise in *h(t)*

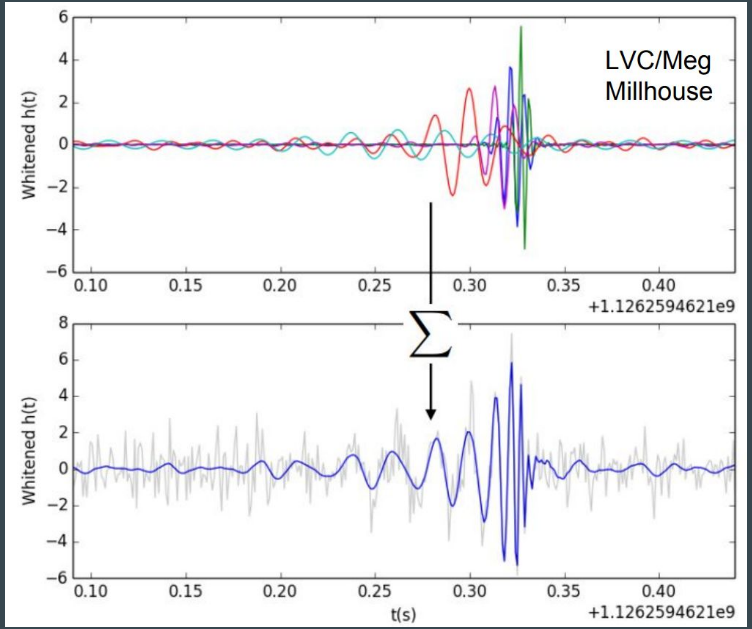


Davis et al. (2022)

[4] Davis et al. (2019)

Noise modelling

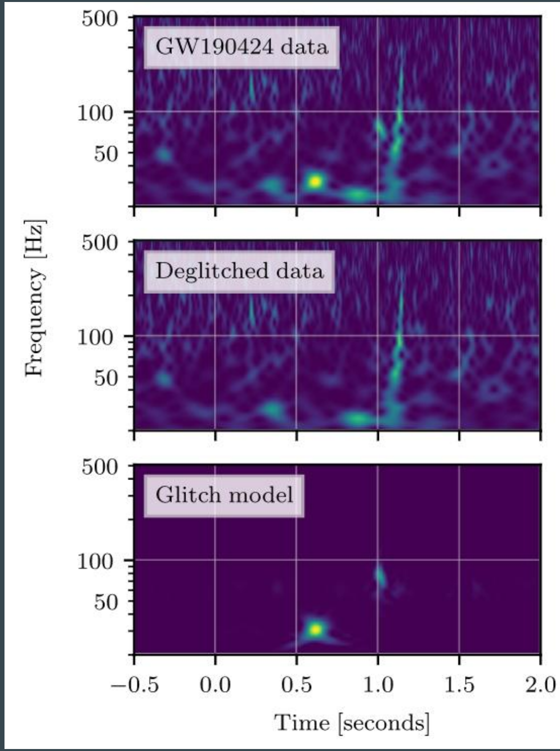
- Quite often the root cause of noise cannot be found
- Instead, try to model the noise
  - Linear noise subtraction[4]
    - Use witness channel to model the noise in *h(t)*
  - BayesWave[5]
    - Models the noise using wavelets (no witness channel information is used)



[4] Davis et al. (2019)
[5] Cornish et al. (2021)

Noise modelling

- Quite often the root cause of noise cannot be found
- Instead, try to model the noise
  - Linear noise subtraction[4]
    - Use witness channel to model the noise in $h(t)$
  - BayesWave[5]
    - Models the noise using wavelets (no witness channel information is used)



Davis et al. (2022)

[4] Davis et al. (2019)
[5] Cornish et al. (2021)

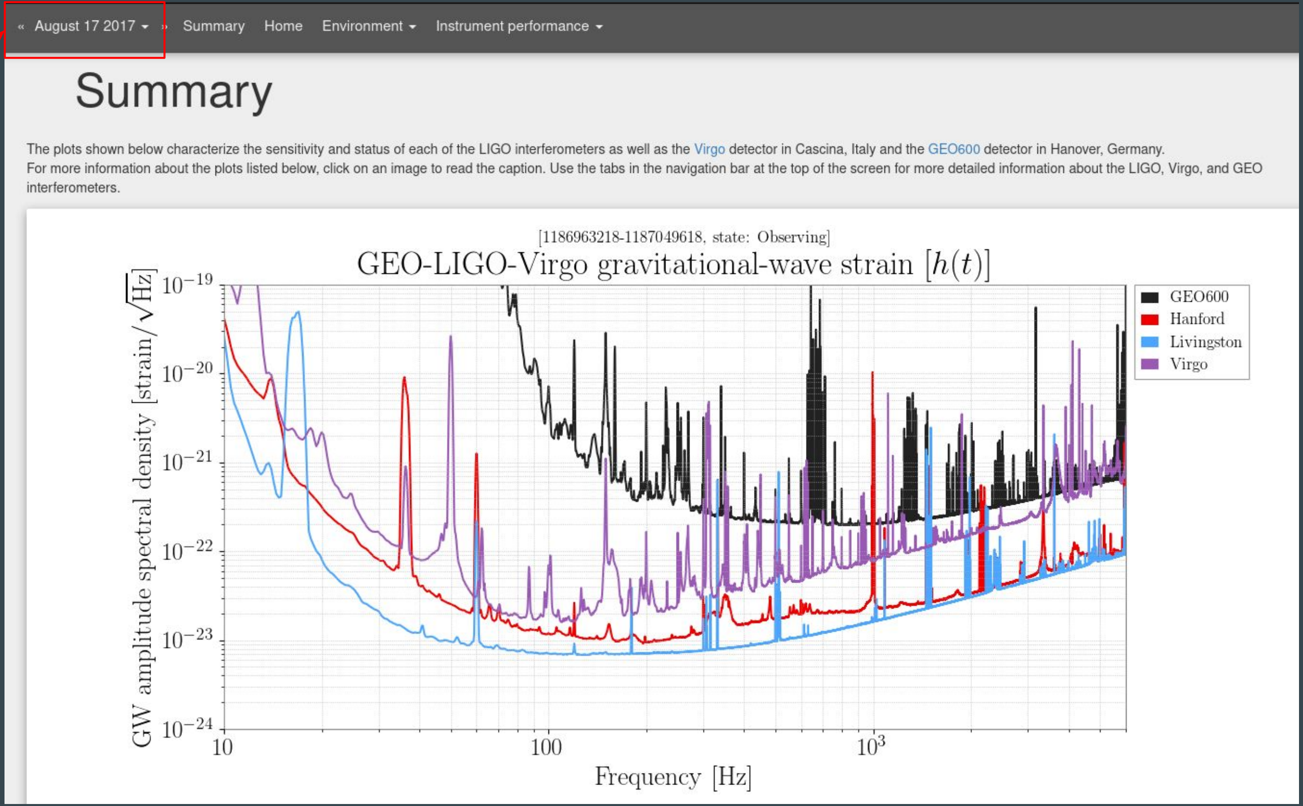
Vetoes

- If nothing else works, we create data quality vetoes
- Different veto categories depending on the severity of the issue
  - Category 1: Major issue with a key detector component
  - Category 2: Known noise coupling to *h(t)*, e.g. high ground motion
  - Category 3: statistical noise coupling to *h(t)* that is not very well understood

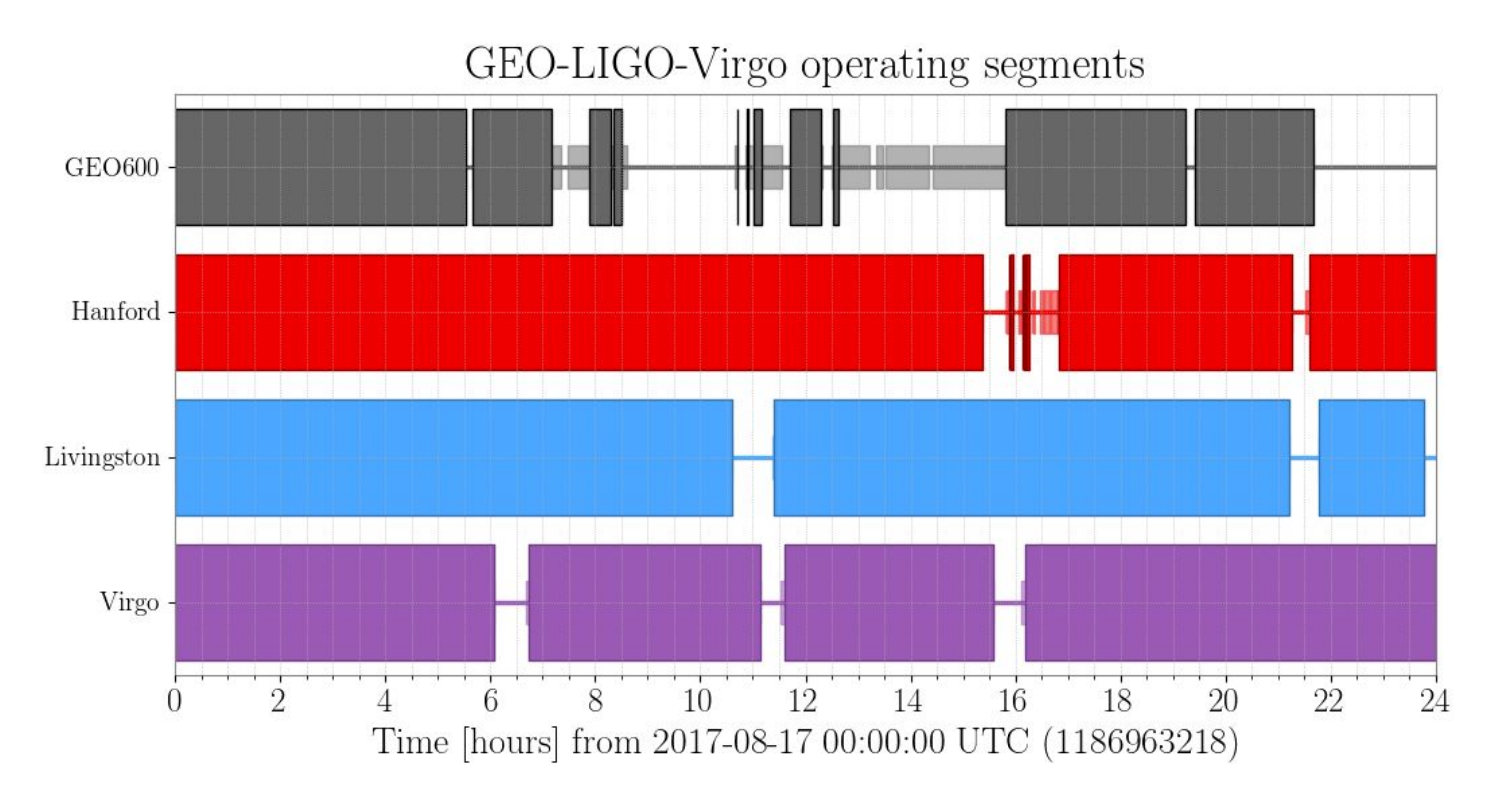
Bit	Short Name	Description
Data Quality Bits		
0	DATA	data present
1	CBC_CAT1	passes the cbc CAT1 test
2	CBC_CAT2	passes cbc CAT2 test
3	CBC_CAT3	passes cbc CAT3 test
4	BURST_CAT1	passes burst CAT1 test
5	BURST_CAT2	passes burst CAT2 test
6	BURST_CAT3	passes burst CAT3 test

Daily detector status (link)

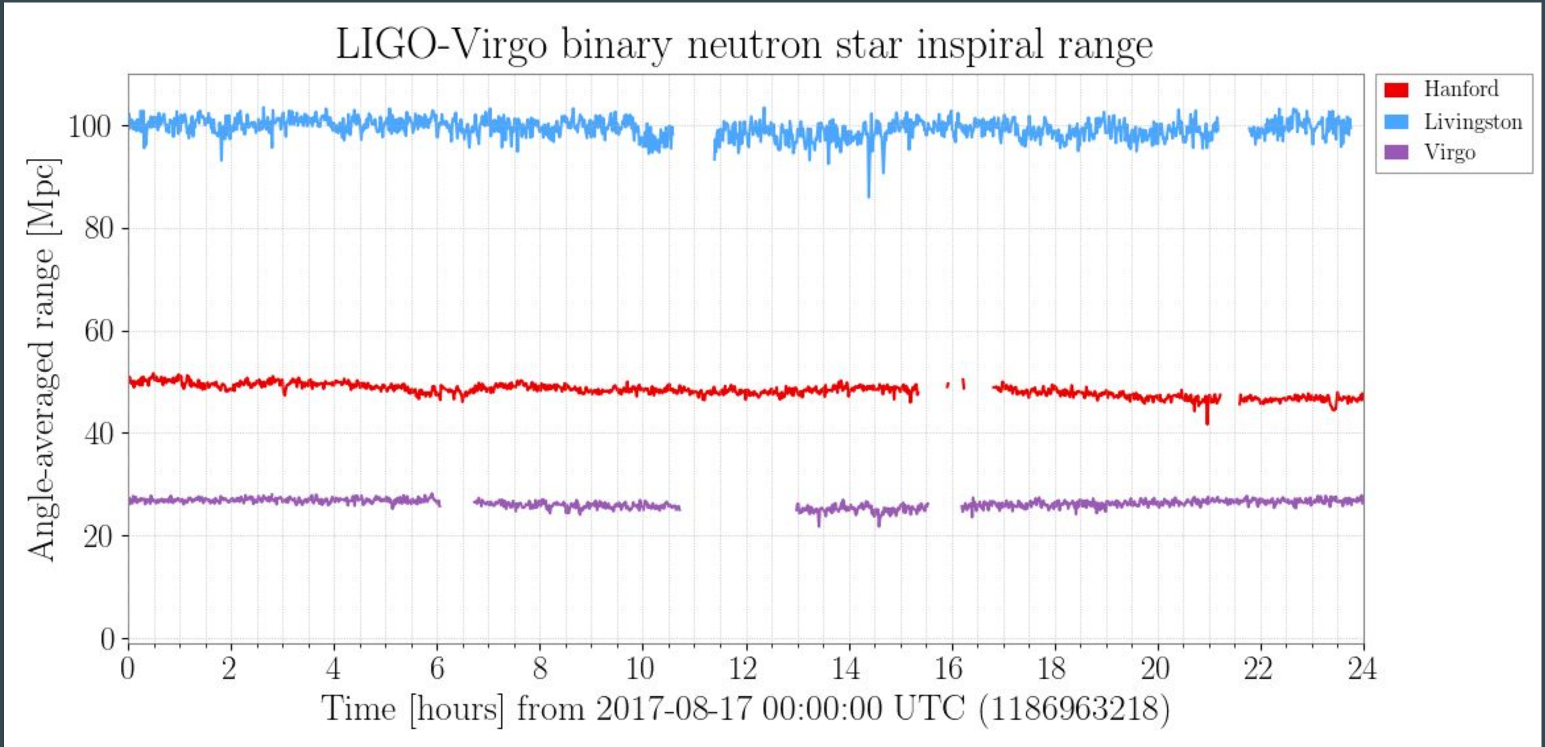
Date selection



Daily detector status (link)

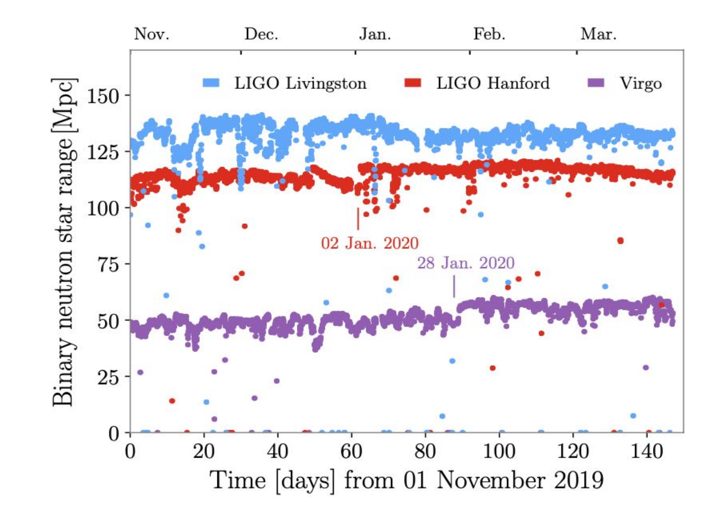


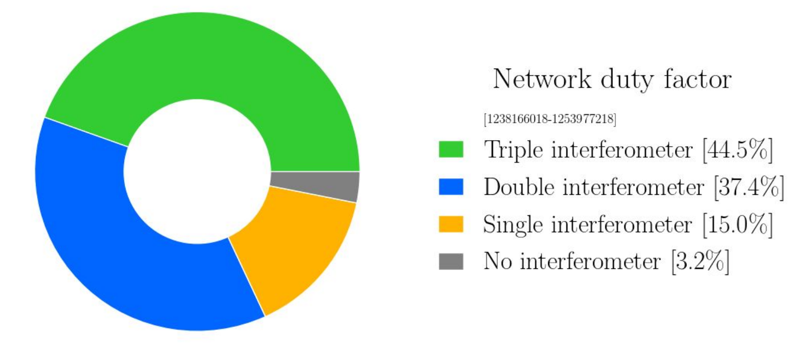
Daily detector status (link)



O3 summary

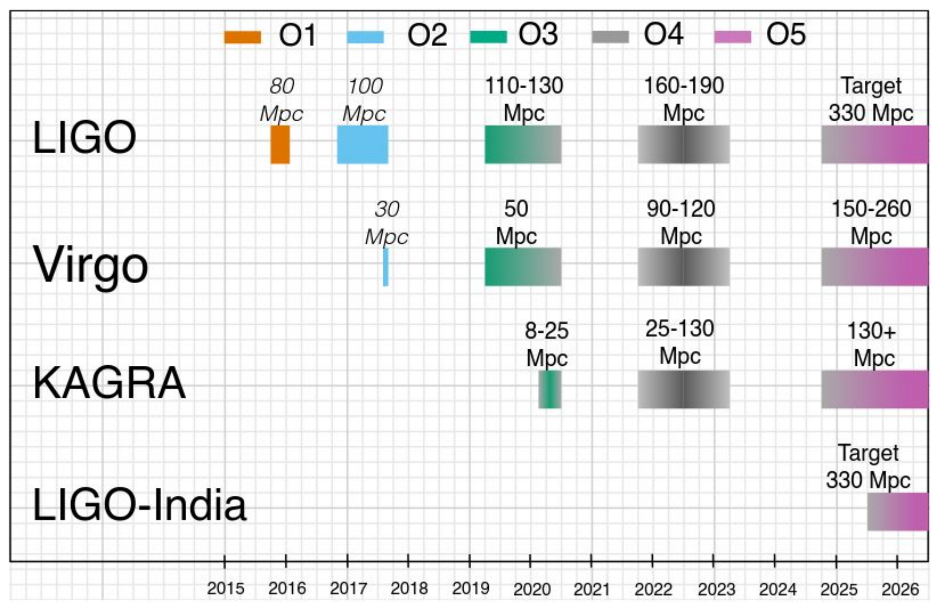
- Split between
  - O3a (Apr 1 2019 - Sept 30 2019)
    - Abbott et al. (2021)
  - O3b (Nov 1 2019 - Mar 27 2020)
    - LIGO, Virgo, KAGRA (2021)
- 74 GWs detected
  - 39 in O3a
  - 35 in O3b
- 18/74 (24%) of O3 GW candidates required glitch mitigation





Plans for O4 (link)

- Observing run 4 is scheduled to start on May 24, 2023
- Fractional increases in sensitivity result in many more detections!
    - (160/130) ≈ 1.2 □ 1.8 more signals
- Changes in interferometers
    - Higher laser power
    - Low noise mitigation
    - New end test mass mirrors
    - Frequency-dependent squeezing
    - ...



Useful data quality references

- LIGO Strain Data
  - A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals
  - Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914
  - The search for gravitational wave bursts in data from the second LIGO science run
- LIGO Data Quality
  - LIGO Data Quality in the Second and Third Observing Runs
  - Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914
  - Sensitivity and performance of the Advanced LIGO detectors in the third observing run
  - Environmental noise in Advanced LIGO detectors
  - Frequency-Dependent Squeezing for Advanced LIGO

Thank you!

Questions?