Data Quality and GWpy

Open Data Workshop #3
May 27, 2020

Marissa Walker

Acknowledgments

Many slides adapted from Jess McIver and Duncan Macleod https://dcc.ligo.org/LIGO-G1800649/public (2018 Open Data Workshop)

Outline

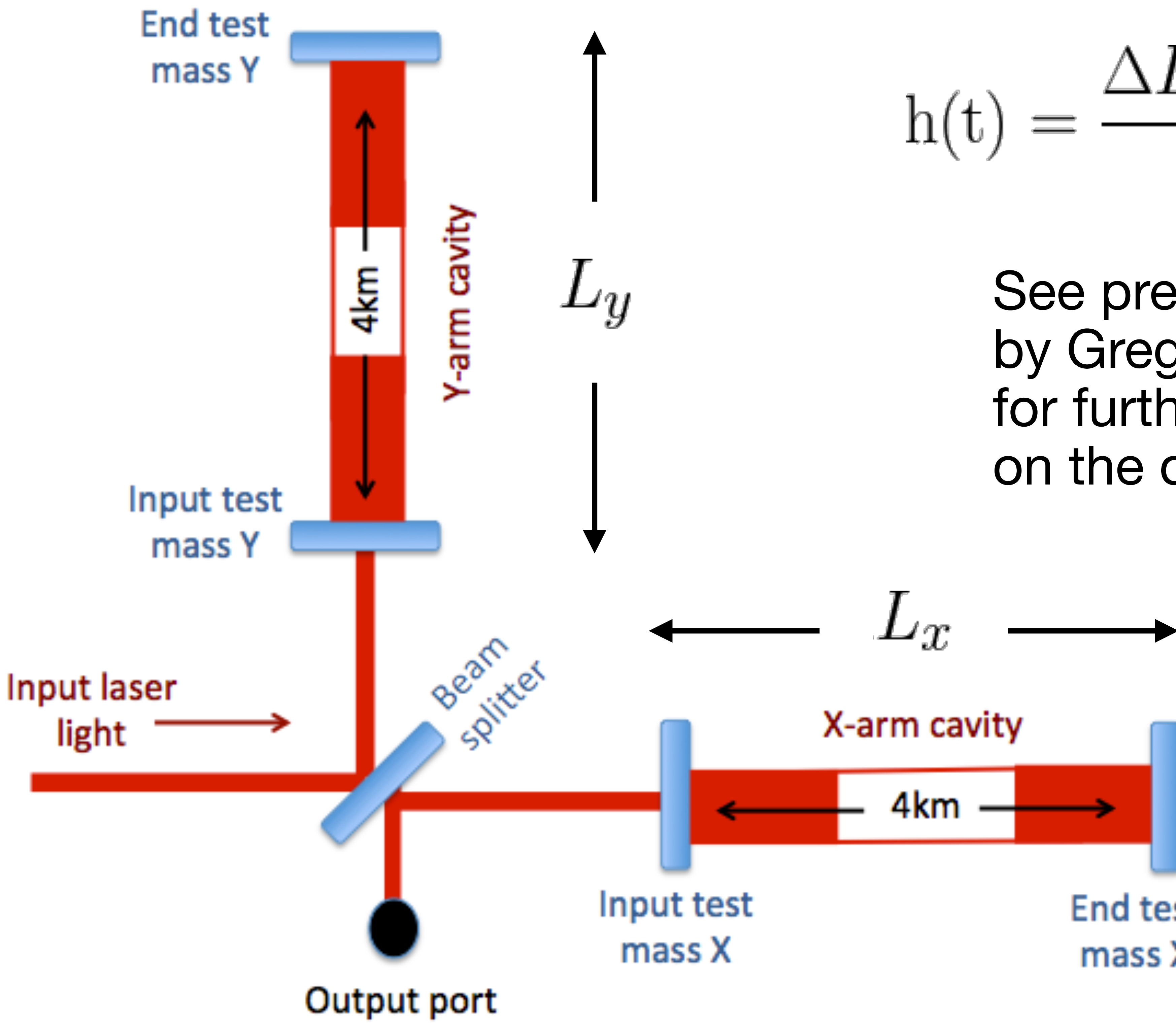
What does LIGO strain data look like?
- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data
- Glitches
- Lines

Data quality information
- Detector status summary pages
- Data quality segments
- Event validation

Summary of resources and references

What is strain data, h(t) ?

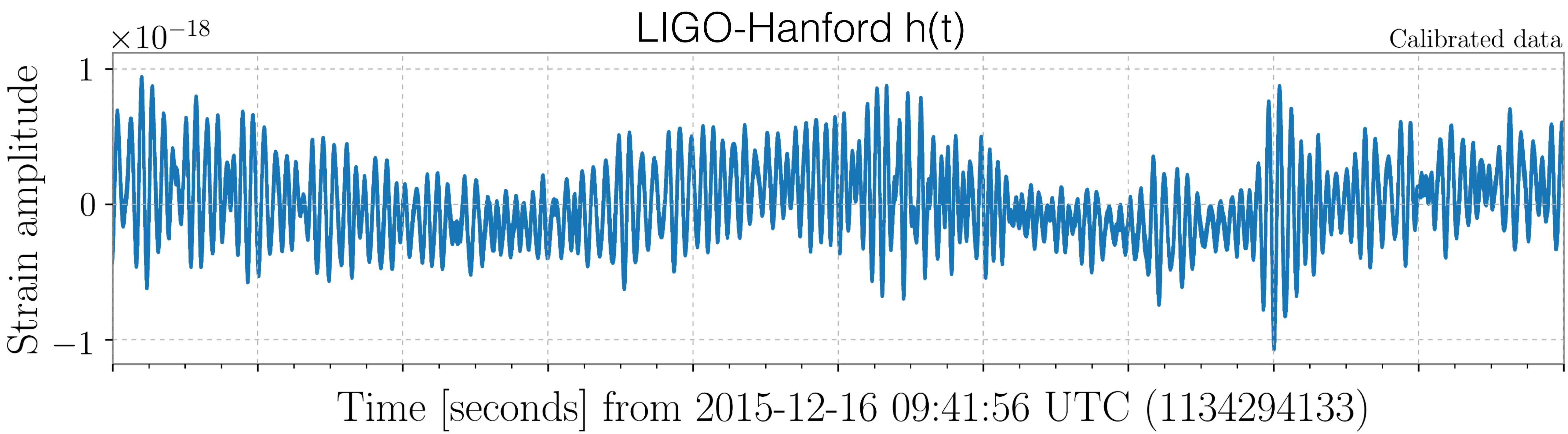


$$\mathrm{h(t)} = \frac{\Delta L_y - \Delta L_x}{L}$$

See previous talk by Gregg Harry for further details on the detectors!

Slide by J McIver

What does LIGO data look like?



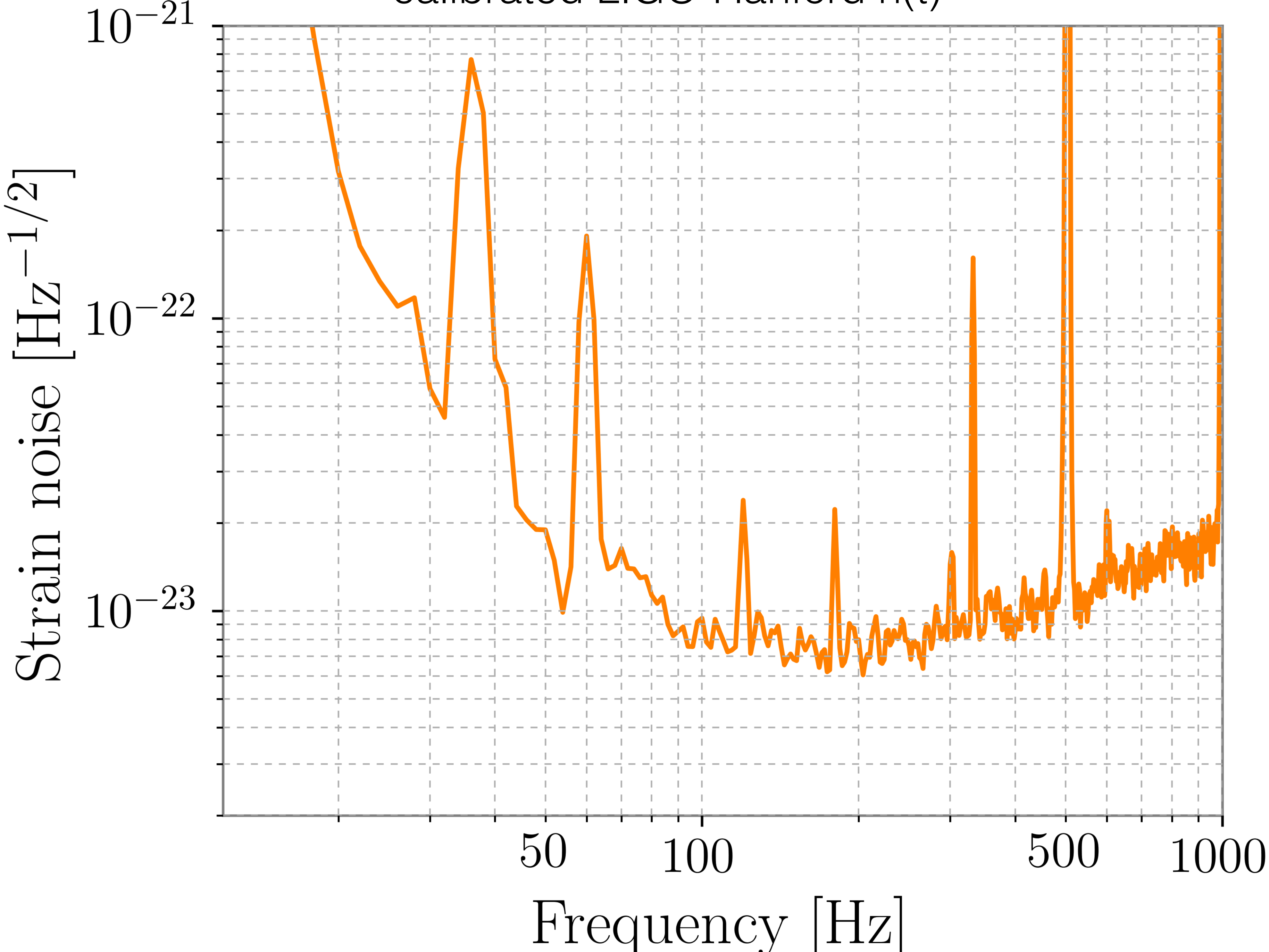
h(t) **sampling rate** for LIGO detectors: 16384 Hz
Open data: 16384 or 4096 Hz

In Tutorials 1.1 and 1.2 you will learn how to get and plot time series data

Made with GWpy by Duncan Macleod

LIGO data in the frequency domain

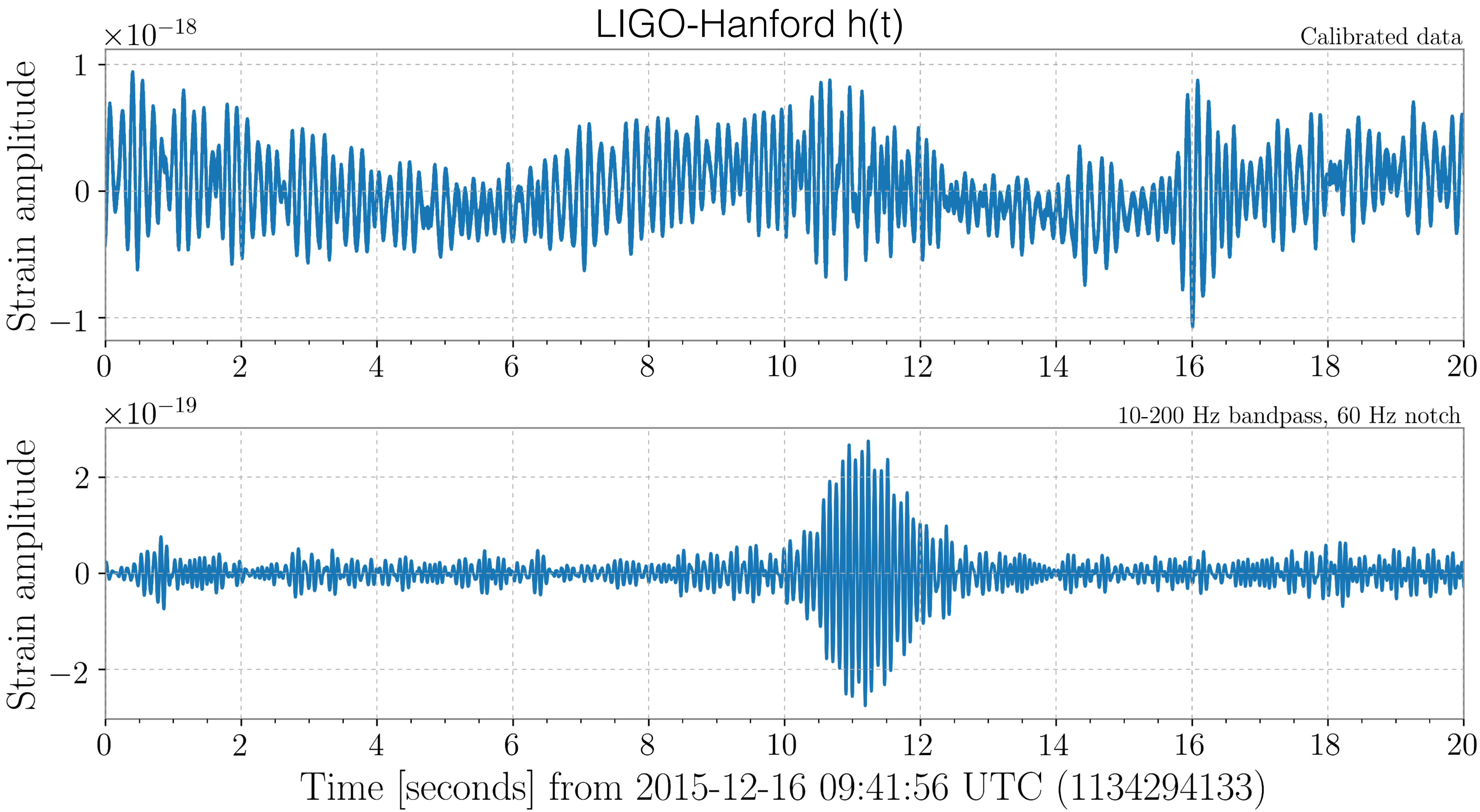
Amplitude Spectral Density (ASD) of calibrated LIGO-Hanford h(t)



In Tutorial 1.2 you will learn how to calculate the amplitude spectral density using GWpy

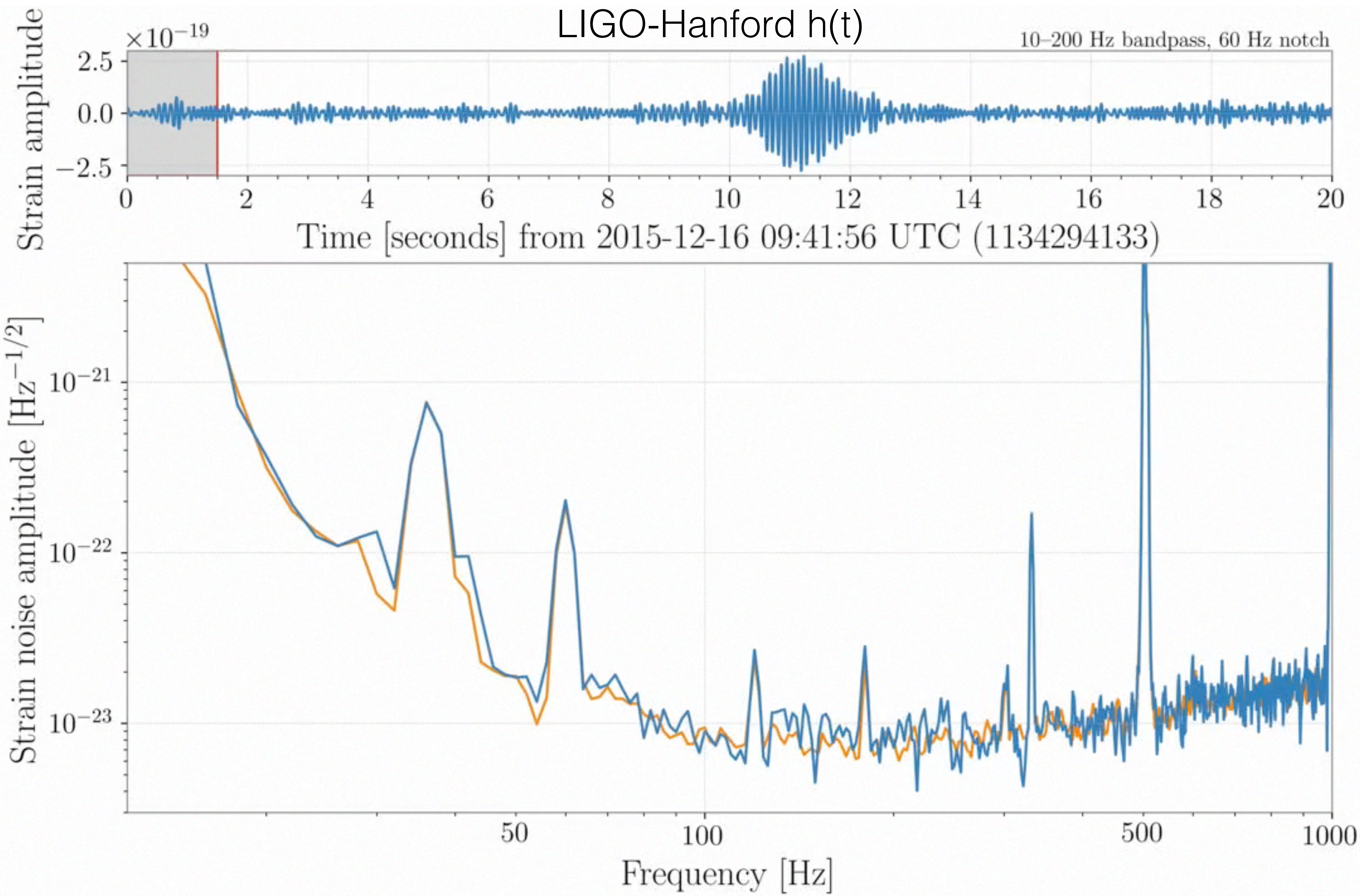
Made with GWpy by Duncan Macleod

What does LIGO data look like?



Made with GWpy by Duncan Macleod

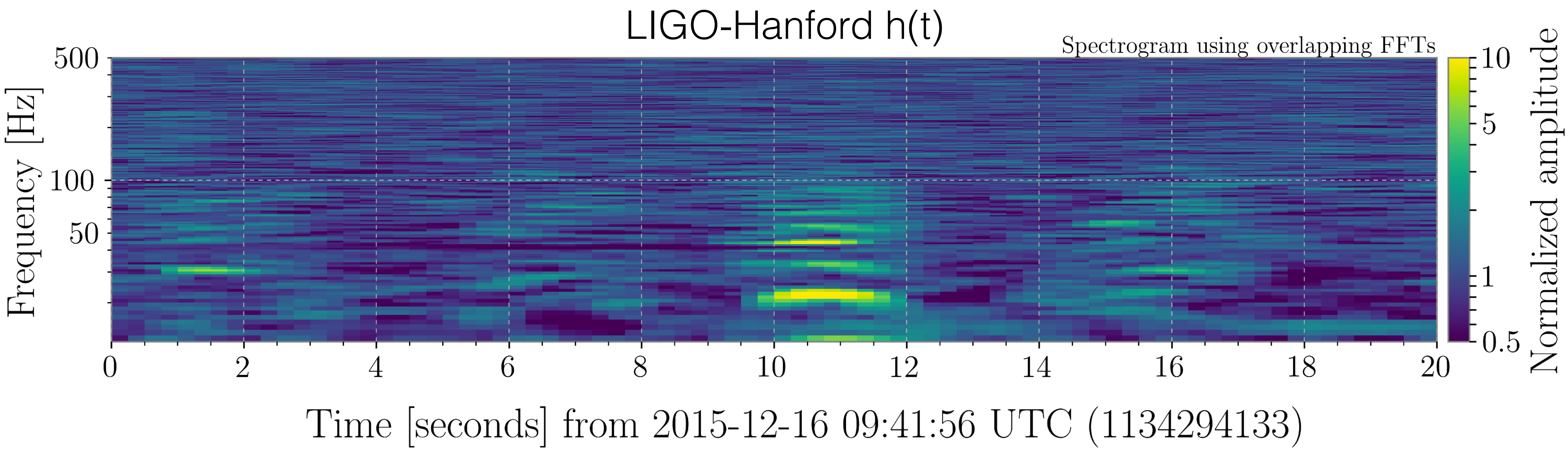
LIGO data in the frequency domain



Made with GWpy by Duncan Macleod. Code: https://git.io/gwpy-ligo-scattering-animation

0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

Time-frequency spectrogram

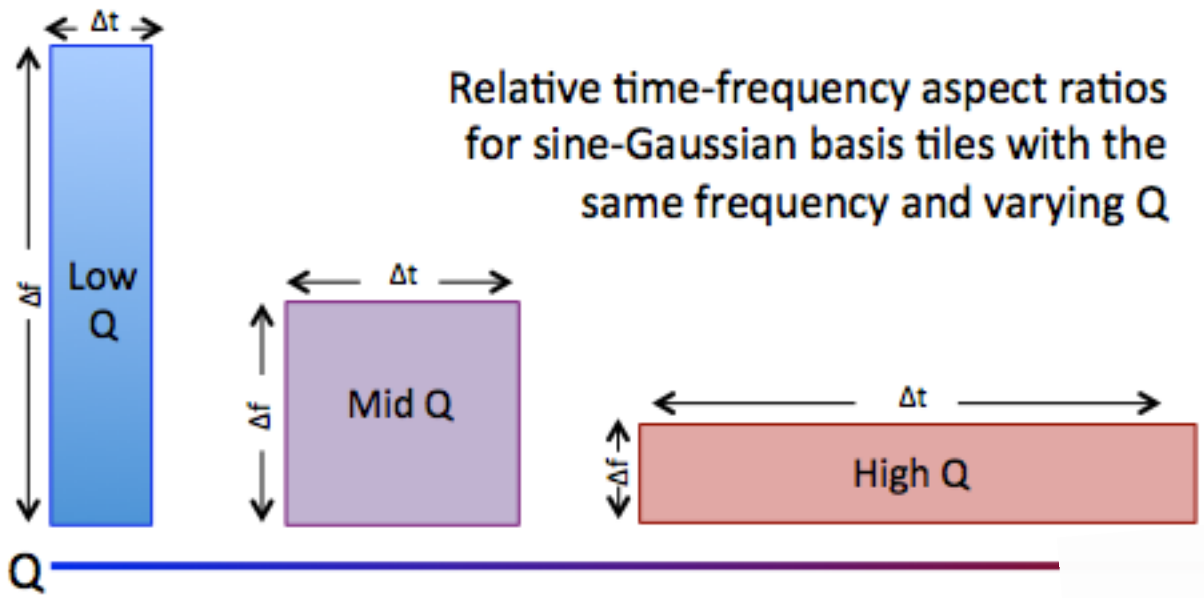


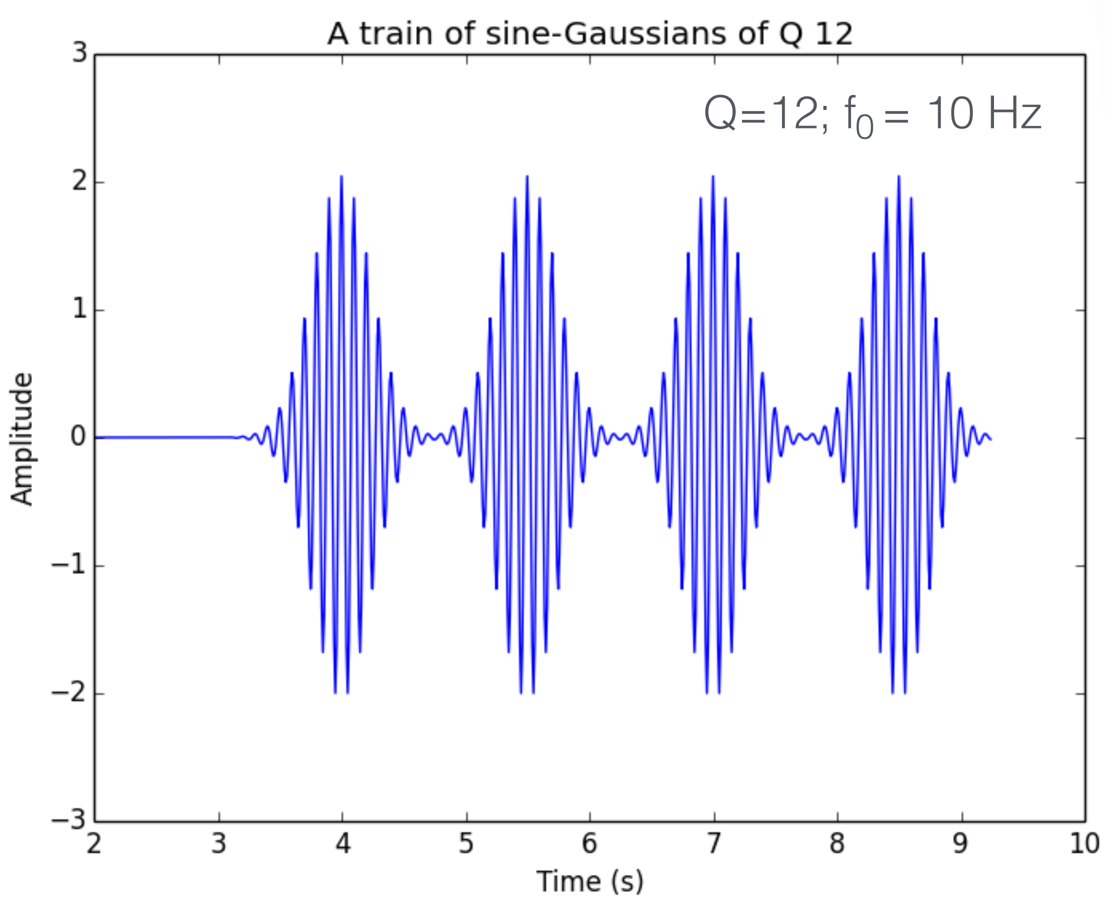
Made with GWpy by Duncan Macleod

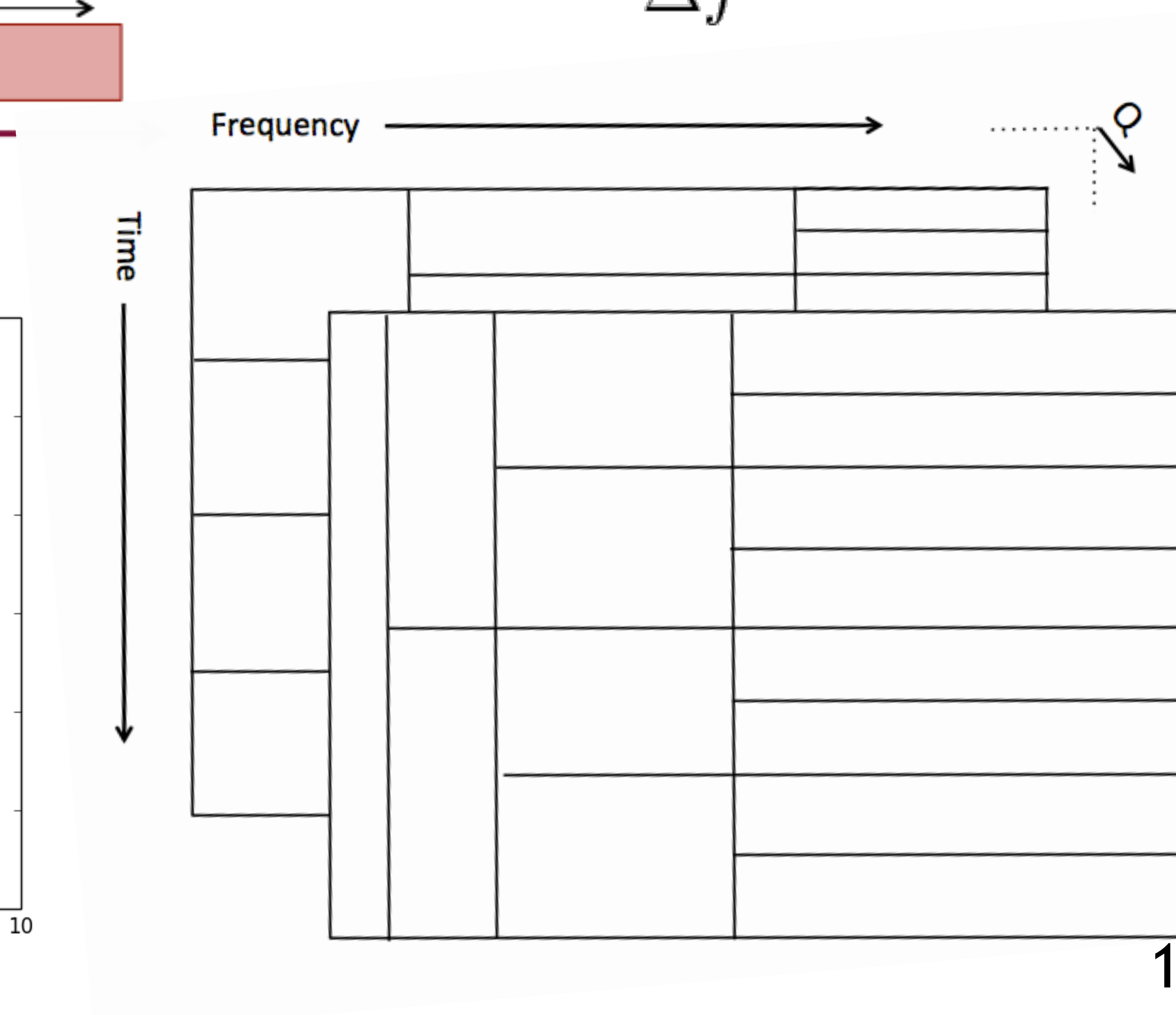
The Q transform

S. Chatterji et al. CQG (2010)
Images: McIver

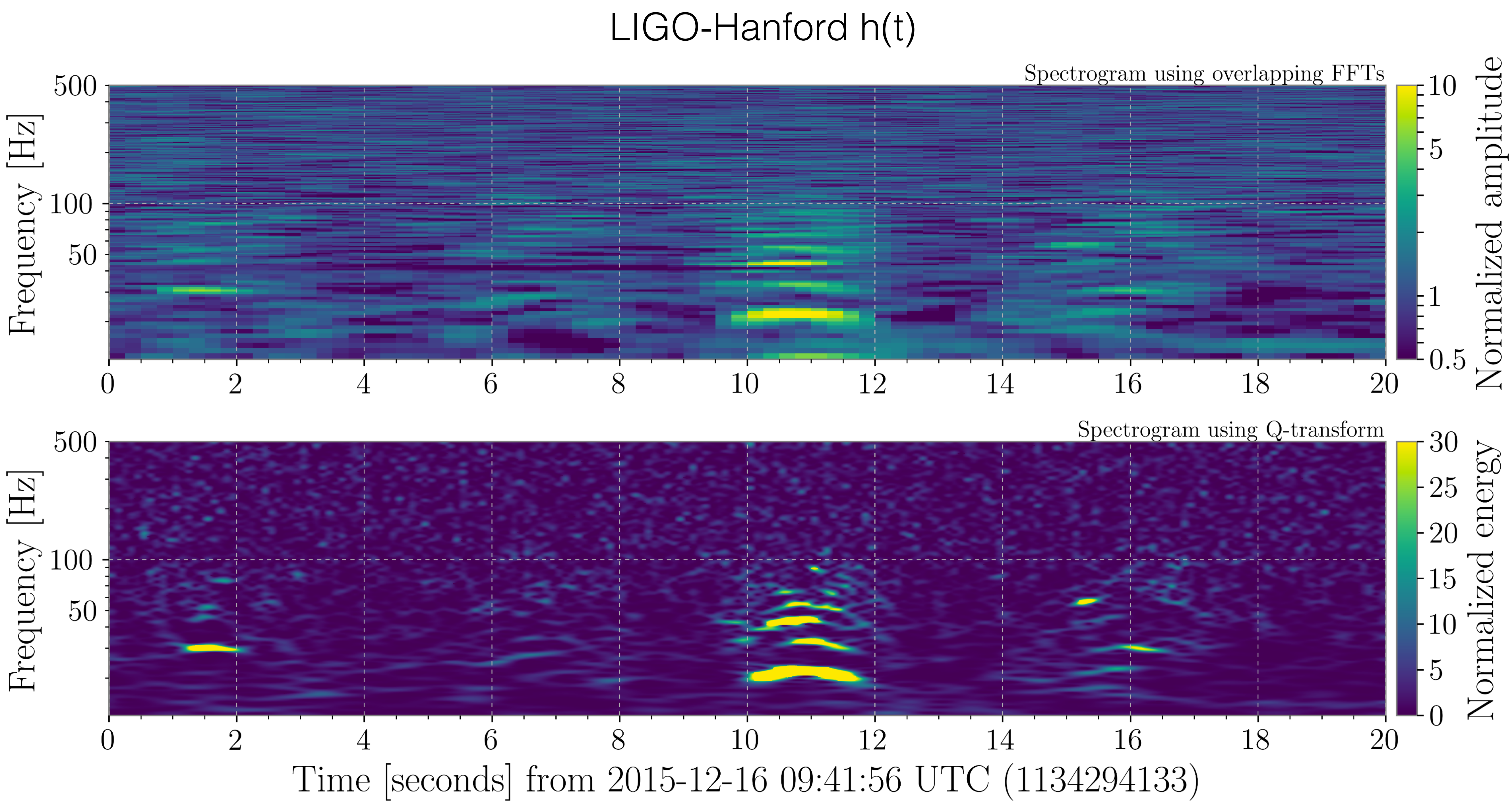
$$Q = \frac{f_0}{\Delta f}$$







Time-frequency spectrograms



In Tutorial 1.3 you will learn how to plot spectrograms and use the q-transform

Made with GWpy by Duncan Macleod

Outline

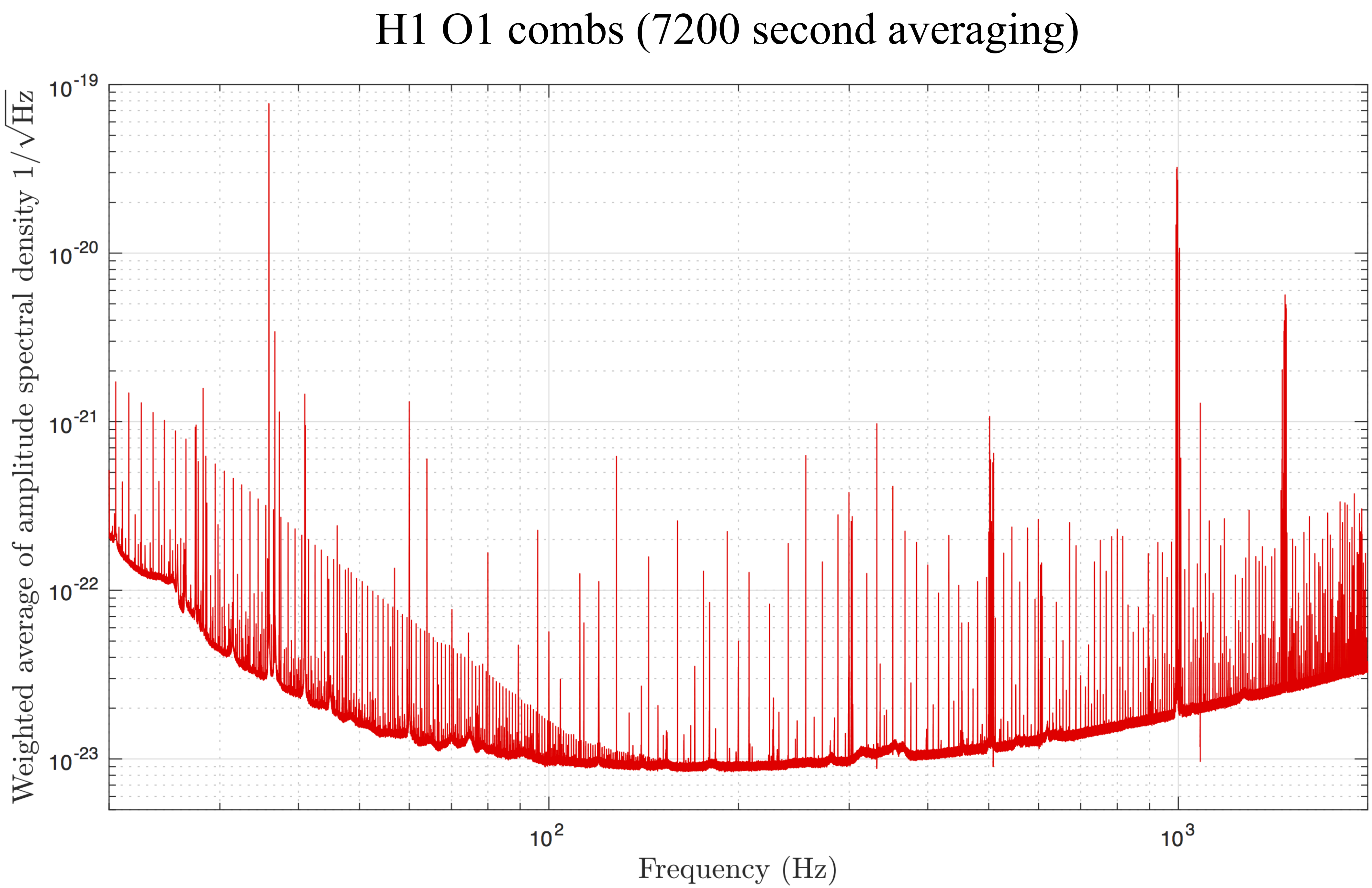
What does LIGO strain data look like?
- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data
- Glitches
- Lines

Data quality information
- Detector status summary pages
- Data quality segments
- Event validation

Summary of resources and references

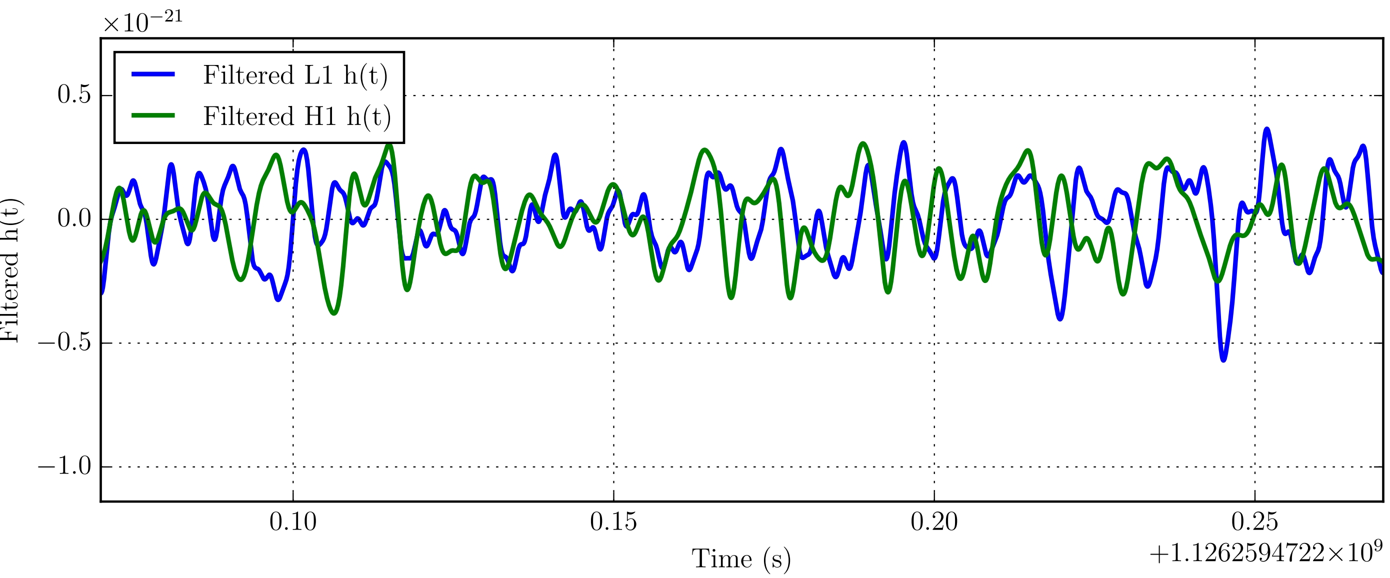
Combs of lines in LIGO data

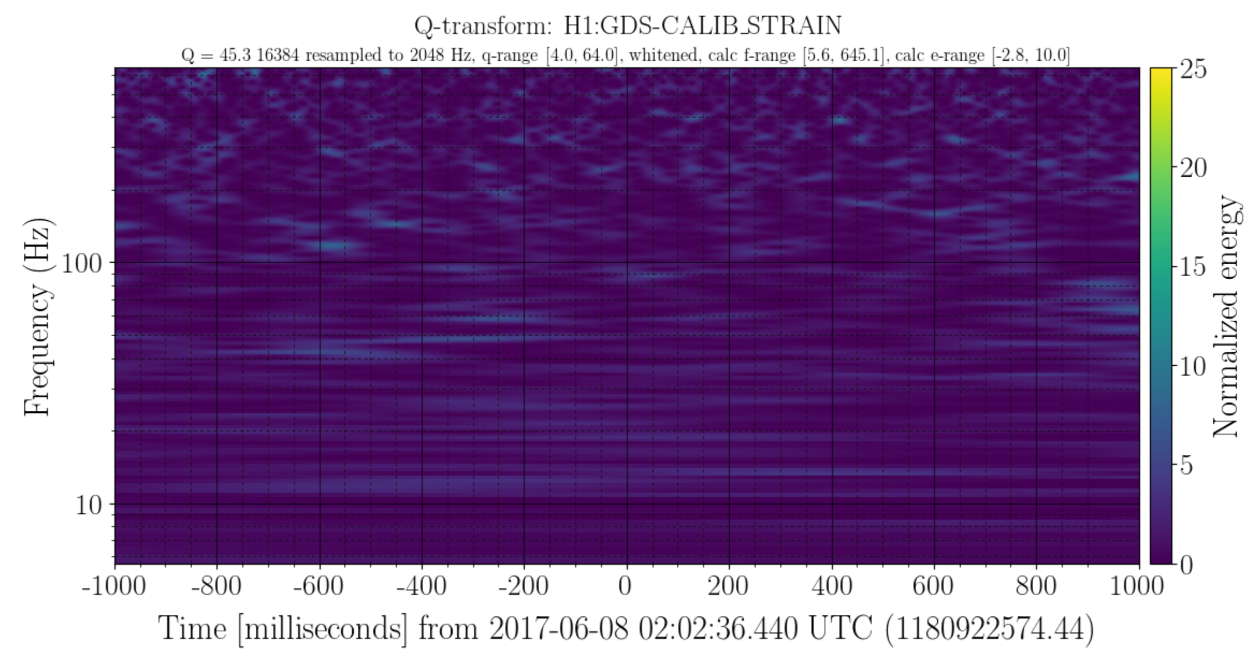


O1 and O2 noise lines paper: Covas et al. (2017) arXiv 1801.07204
O2 Instrumental lines:
https://www.gw-openscience.org/o2speclines/

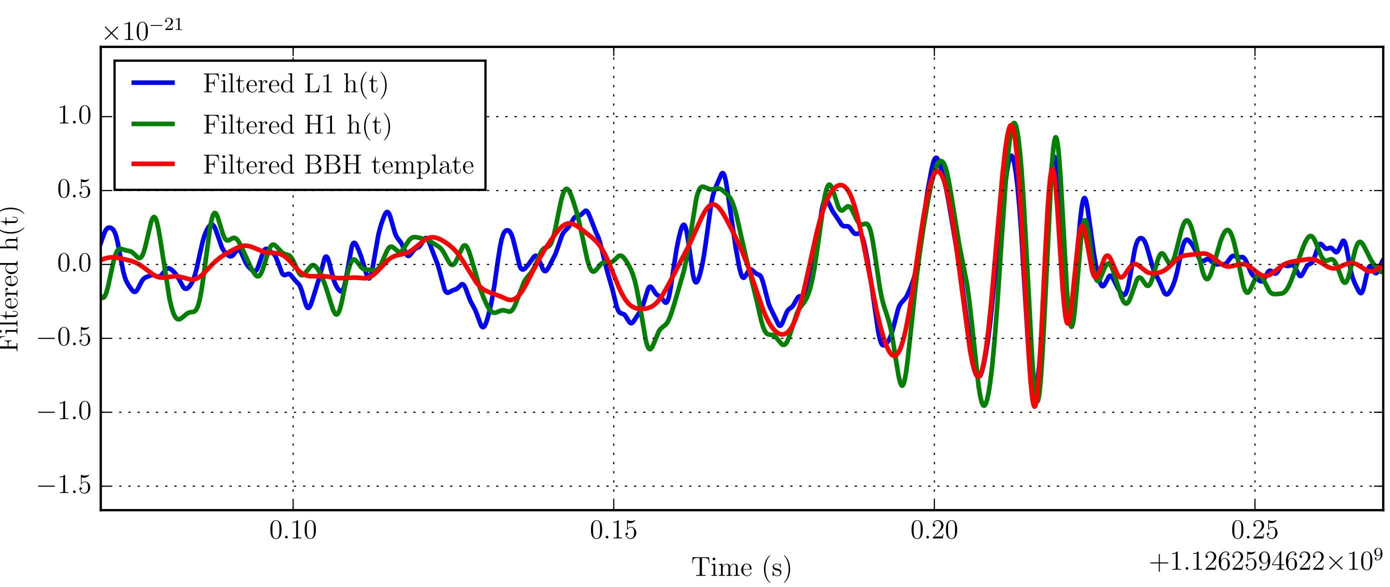
GW Detector data in a perfect world

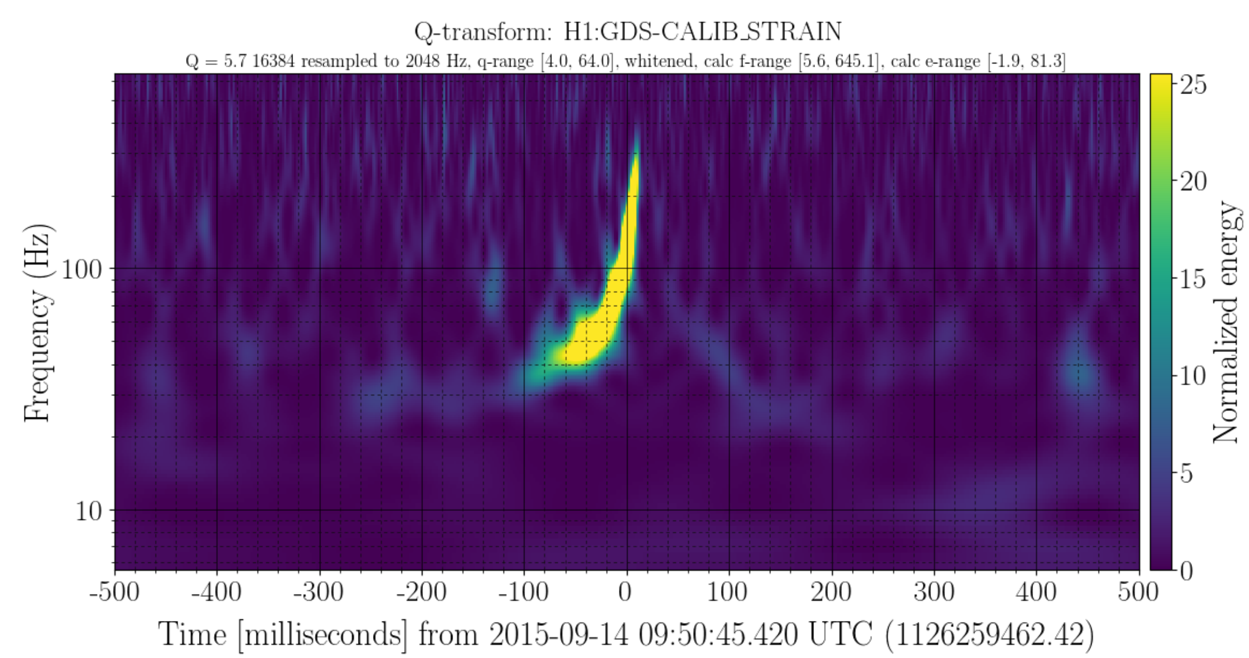
No signal



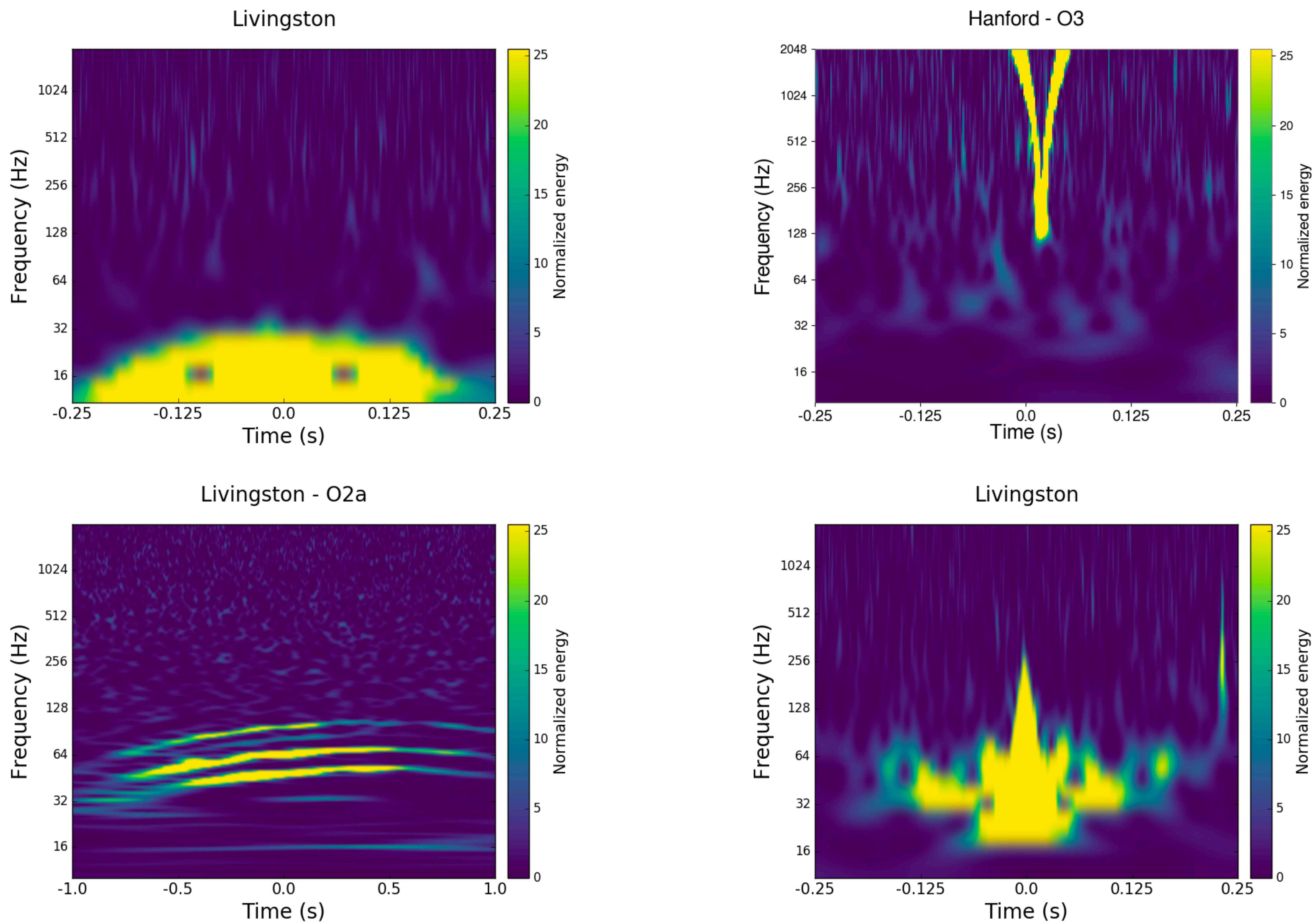


Signal





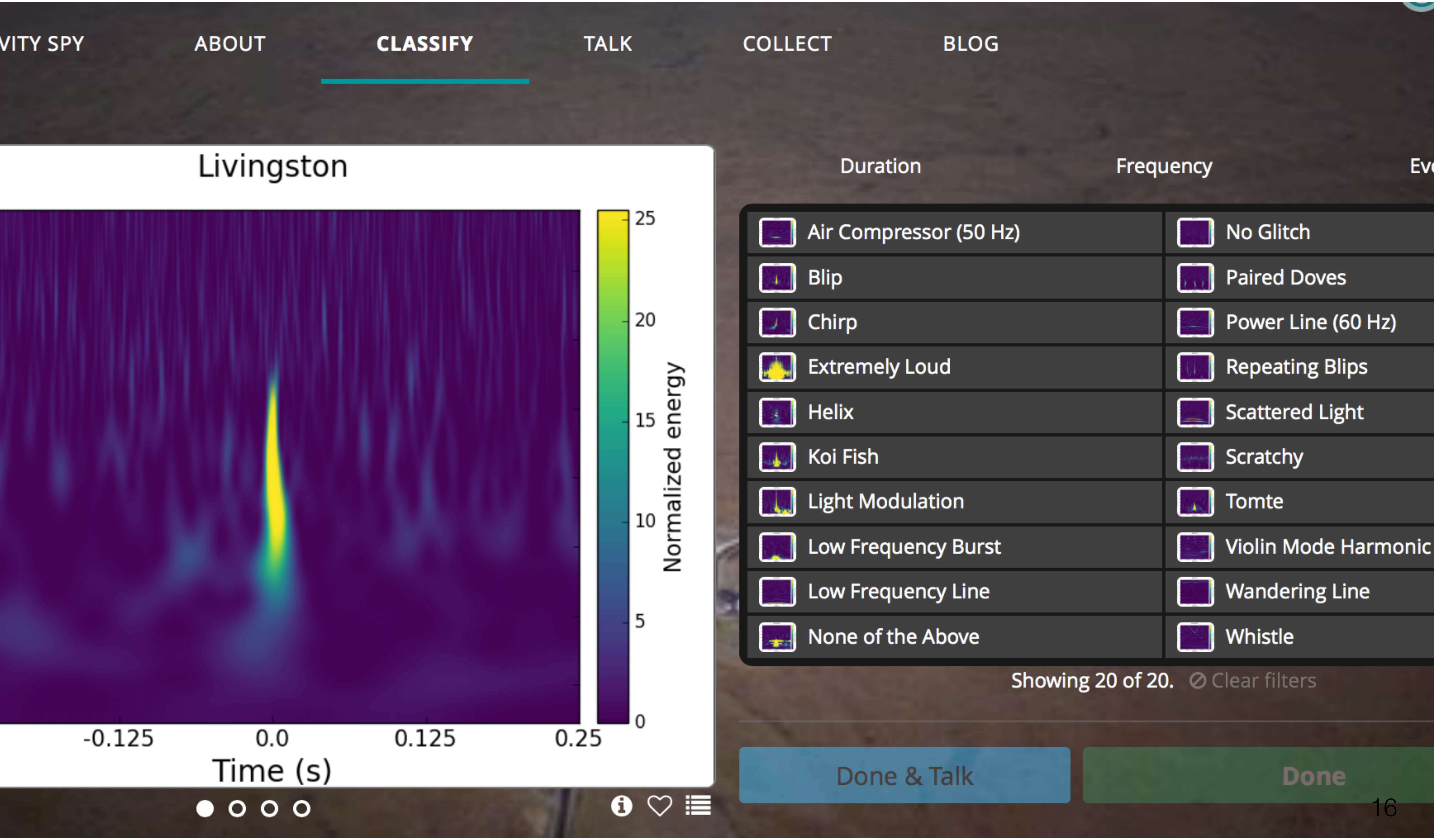
But in the real world we also have…



gravityspy.org

Common glitch types

gravityspy.org   Zevin et al, 2017, CQG



Outline

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Data quality information

- Detector status summary pages
- Data quality segments
- Event validation

Summary of resources and references

Daily detector status (available for O2 & O3)

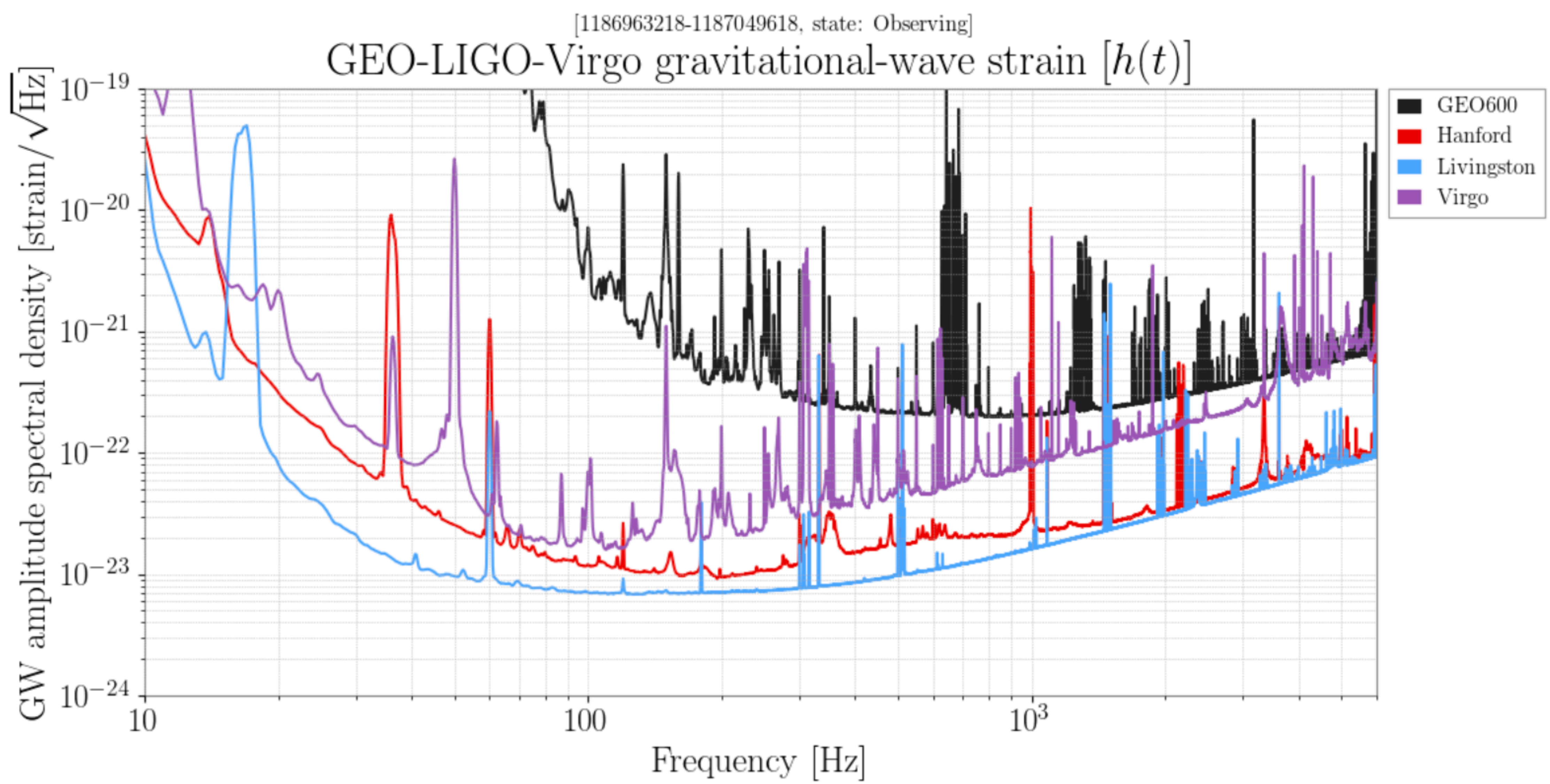
https://www.gw-openscience.org/detector_status/day/20170817/

« August 17 2017 ▾ » Summary Home Environment ▾ Instrument performance ▾

Summary

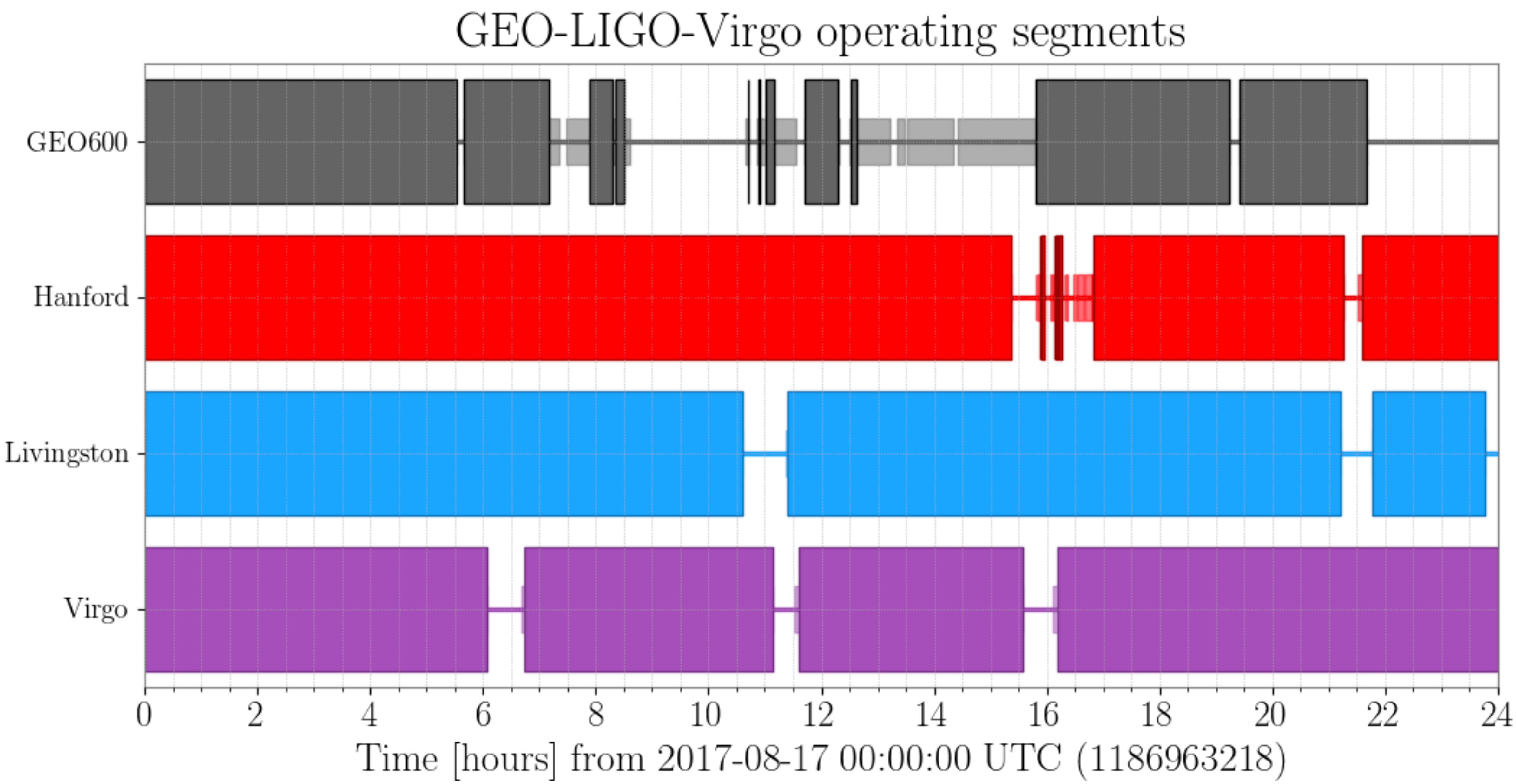
Date selection

The plots shown below characterize the sensitivity and status of each of the LIGO interferometers as well as the Virgo detector in Cascina, Italy and the GEO600 detector in Hanover, Germany.
For more information about the plots listed below, click on an image to read the caption. Use the tabs in the navigation bar at the top of the screen for more detailed information about the LIGO, Virgo, and GEO interferometers.



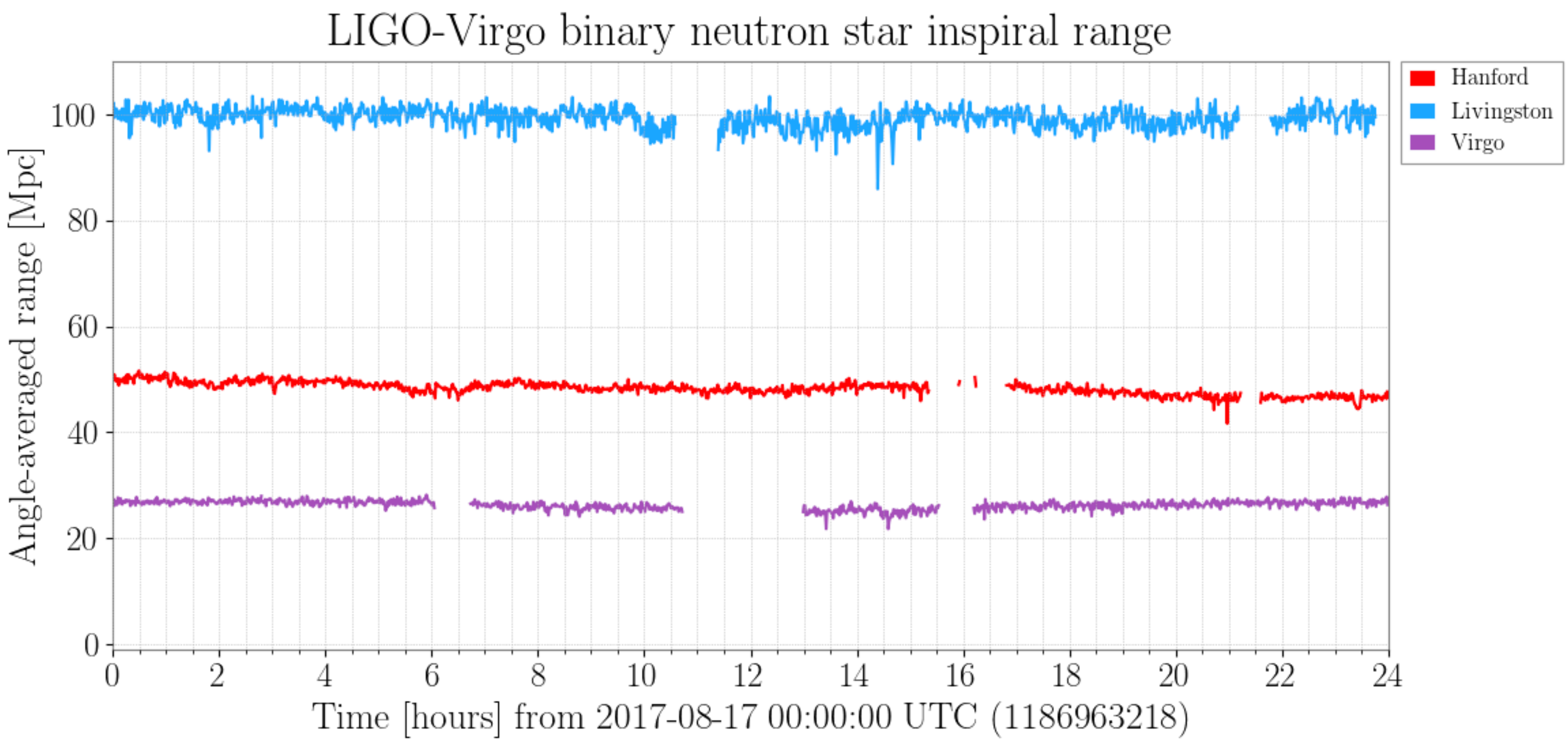
Daily detector status (available for O2 & O3)

https://www.gw-openscience.org/detector_status/day/20170817/



Daily detector status (available for O2 & O3)

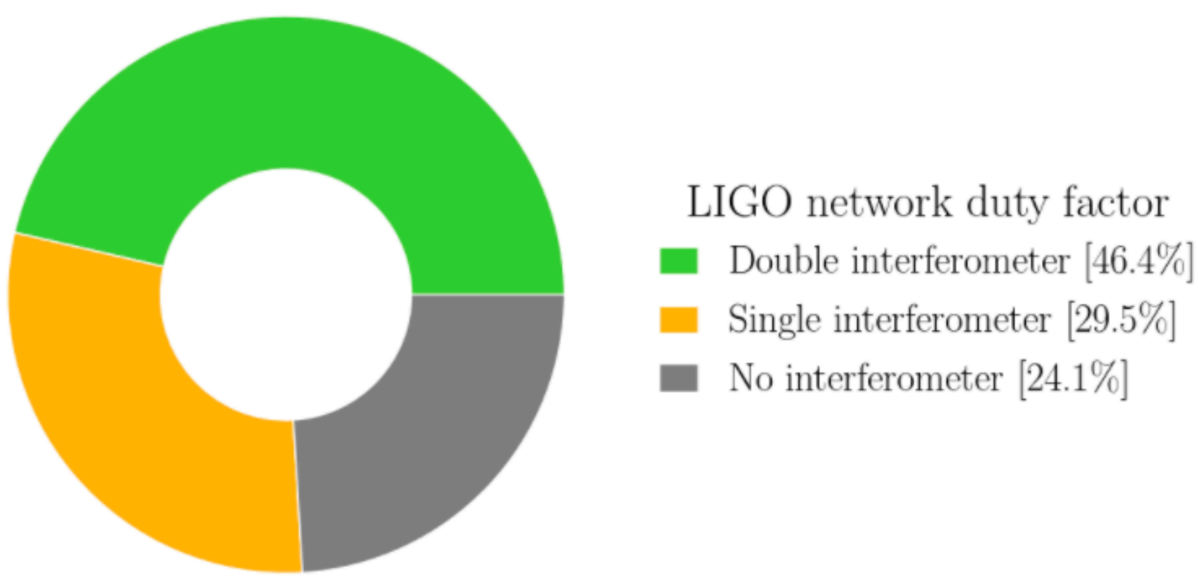
https://www.gw-openscience.org/detector_status/day/20170817/

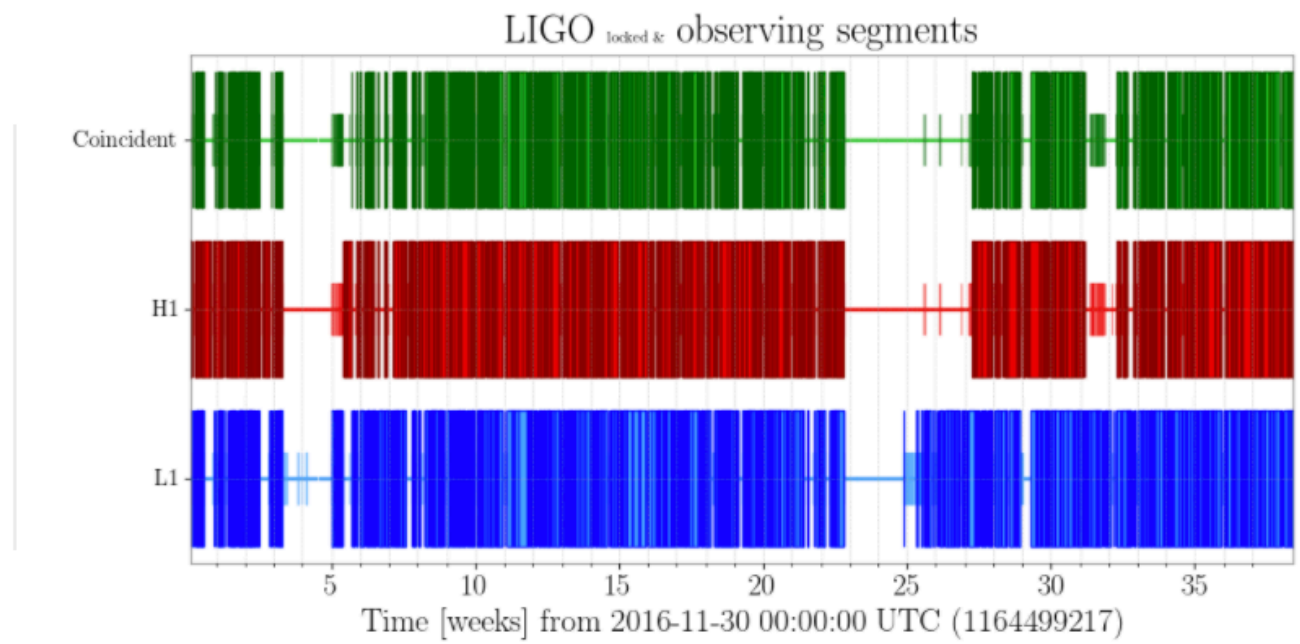


Observing run summaries (O1 or O2)

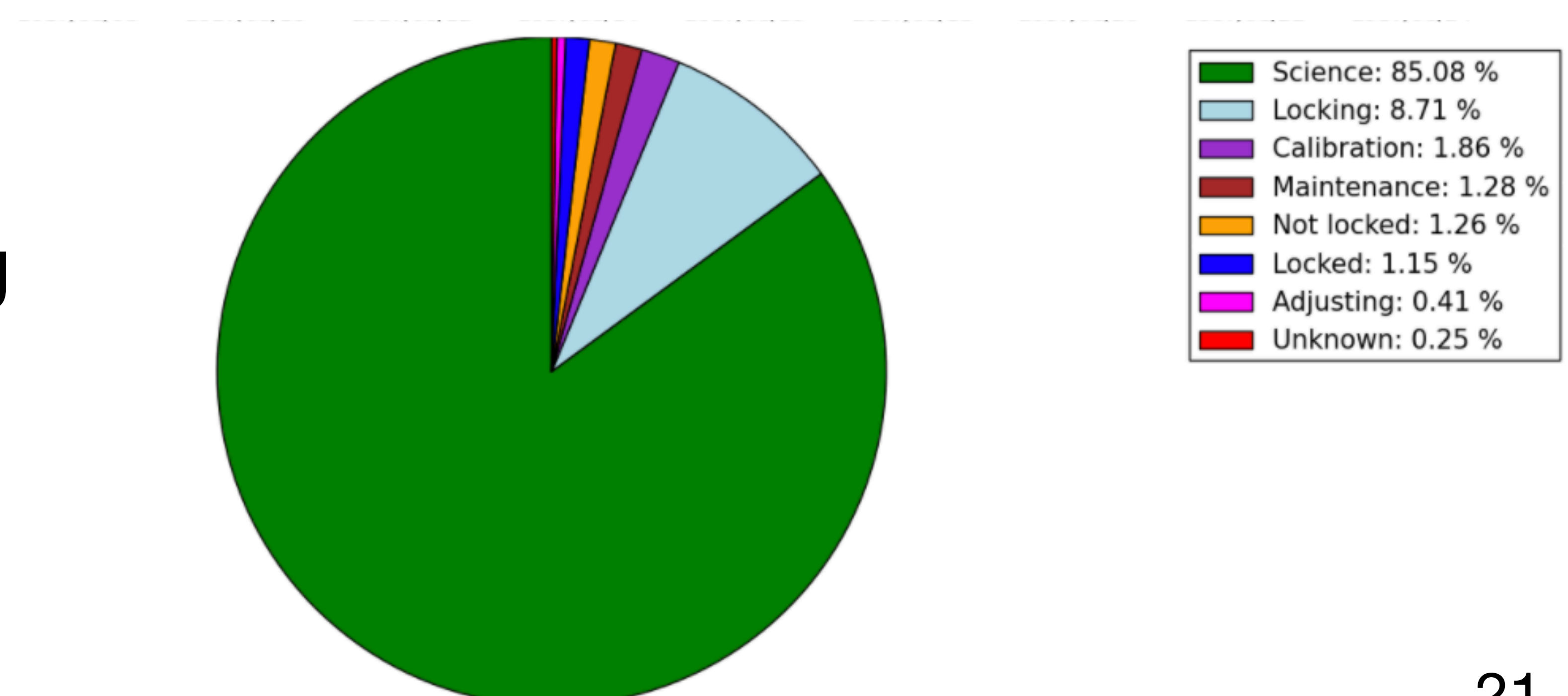
https://www.gw-openscience.org/detector_status/O2/

Includes summary plots of LIGO segments and sensitivity over the run





Also linked is a page showing Virgo's status in O2 (http://www.virgo-gw.eu/O2.html):



Data quality information

DATA (Data Available): Failing this level indicates that LIGO data are not publicly available because the instruments or data calibration were not operating in an acceptable condition.

CAT1 (Category 1): Failing a data quality check at this category indicates **a critical issue with a key detector component not operating in its nominal configuration**.

- These times are identical for each data analysis group.
- *Times that fail CAT1 flags are not available as LIGO open data.*

CAT2 (Category 2): Failing a data quality check at this category indicates times when there is a **known, understood physical coupling to the gravitational wave channel**. For example, high seismic activity.

CAT3 (Category 3): Failing a data quality check at this category indicates times when there is **statistical coupling to the gravitational wave channel** which is not fully understood.

Data quality levels are defined in a cumulative way: a time which fails a given category automatically fails all higher categories.

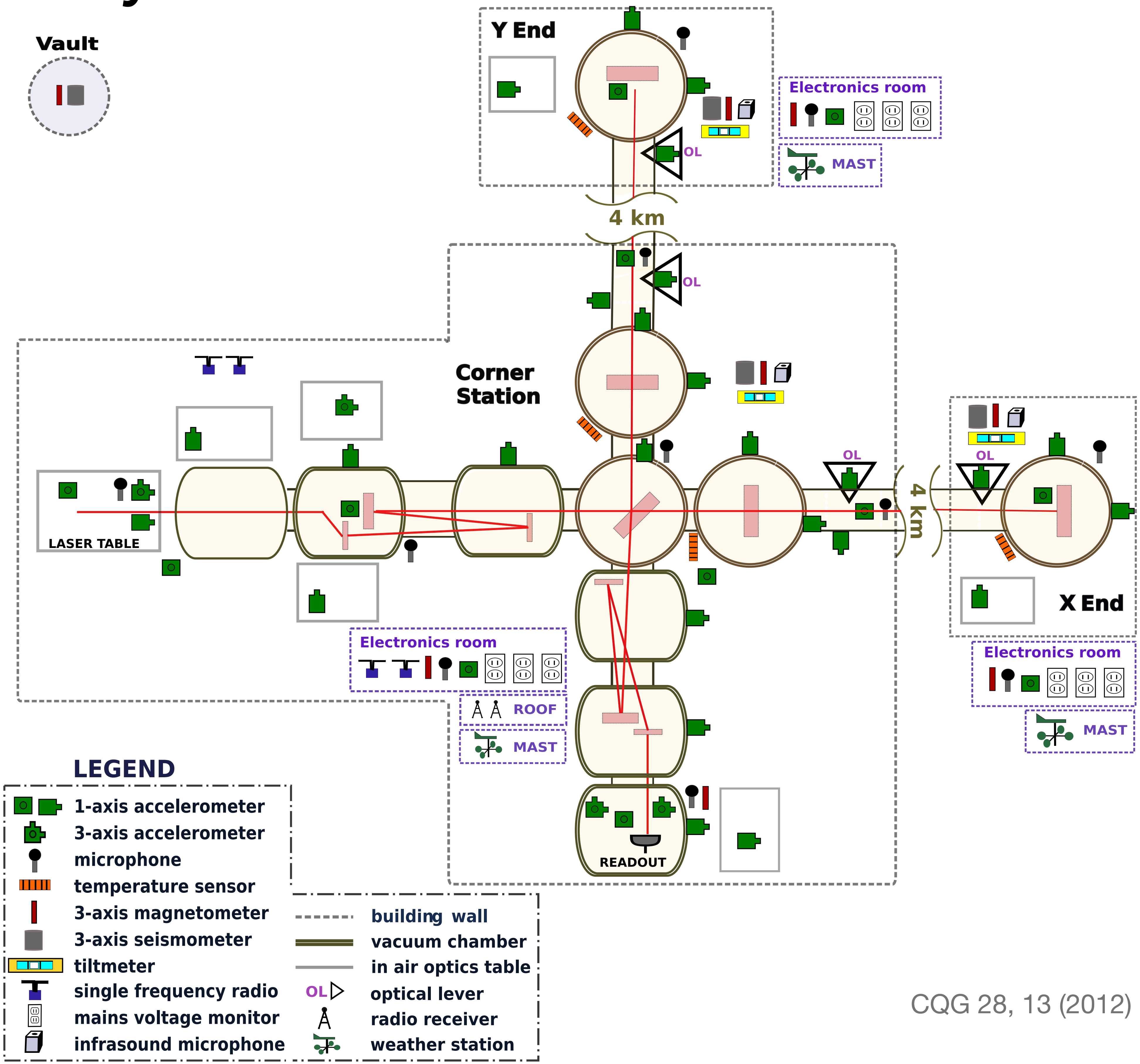
Data quality categories are defined independently for different analysis groups: if something fails at CAT2_BURST, it could pass CAT2_CBC.

Data quality information

Available via the GWOSC

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
Injection Bits		
0	NO_CBC_HW_INJ	no cbc injection
1	NO_BURST_HW_INJ	no burst injections
2	NO_DETCHAR_HW_INJ	no detchar injections
3	NO_CW_HW_INJ	no continuous wave injections
4	NO_STOCH_HW_INJ	no stoch injections

Physical environment channels

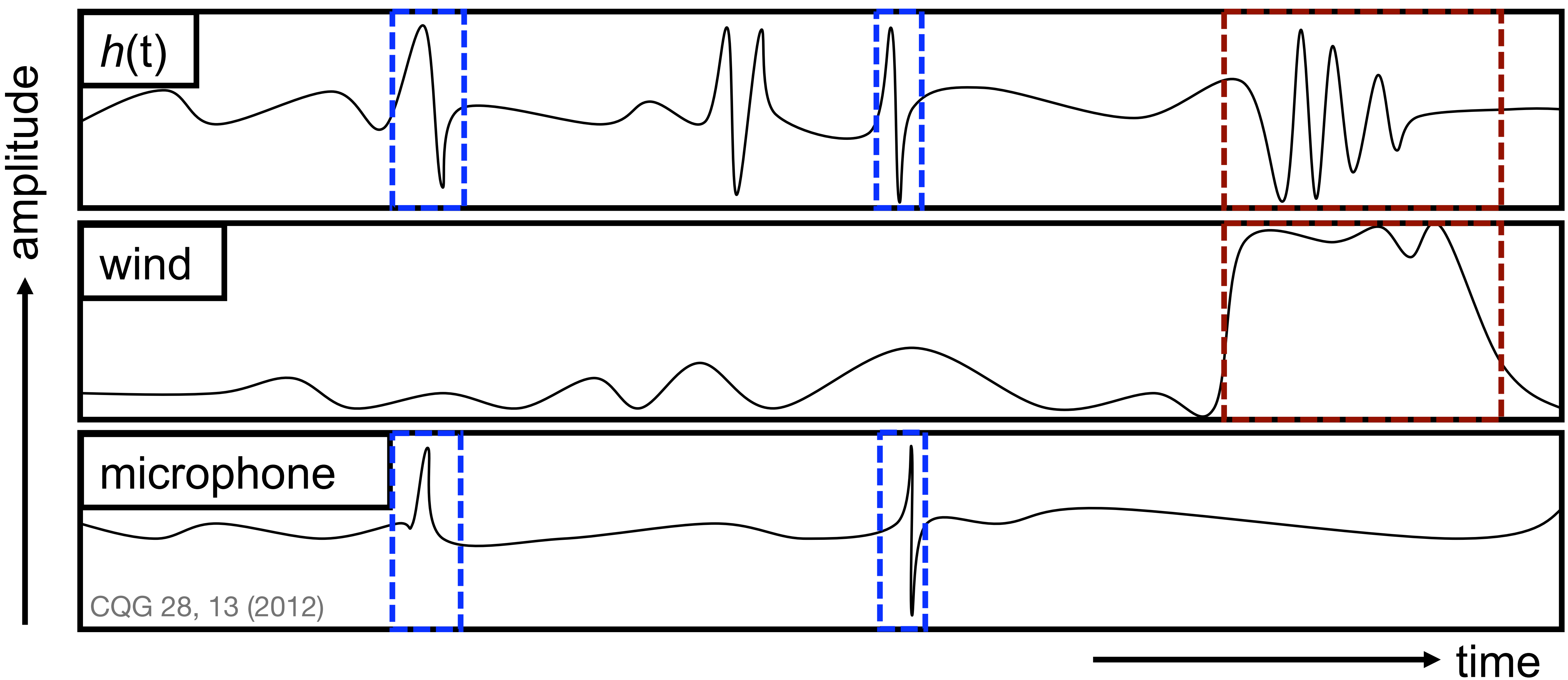


CQG 28, 13 (2012)

Auxiliary channels

We record **over 200,000 channels per detector** that monitor the environment and detector behavior.

We can use these to **help trace the instrumental causes of glitches** that pollute the search backgrounds.

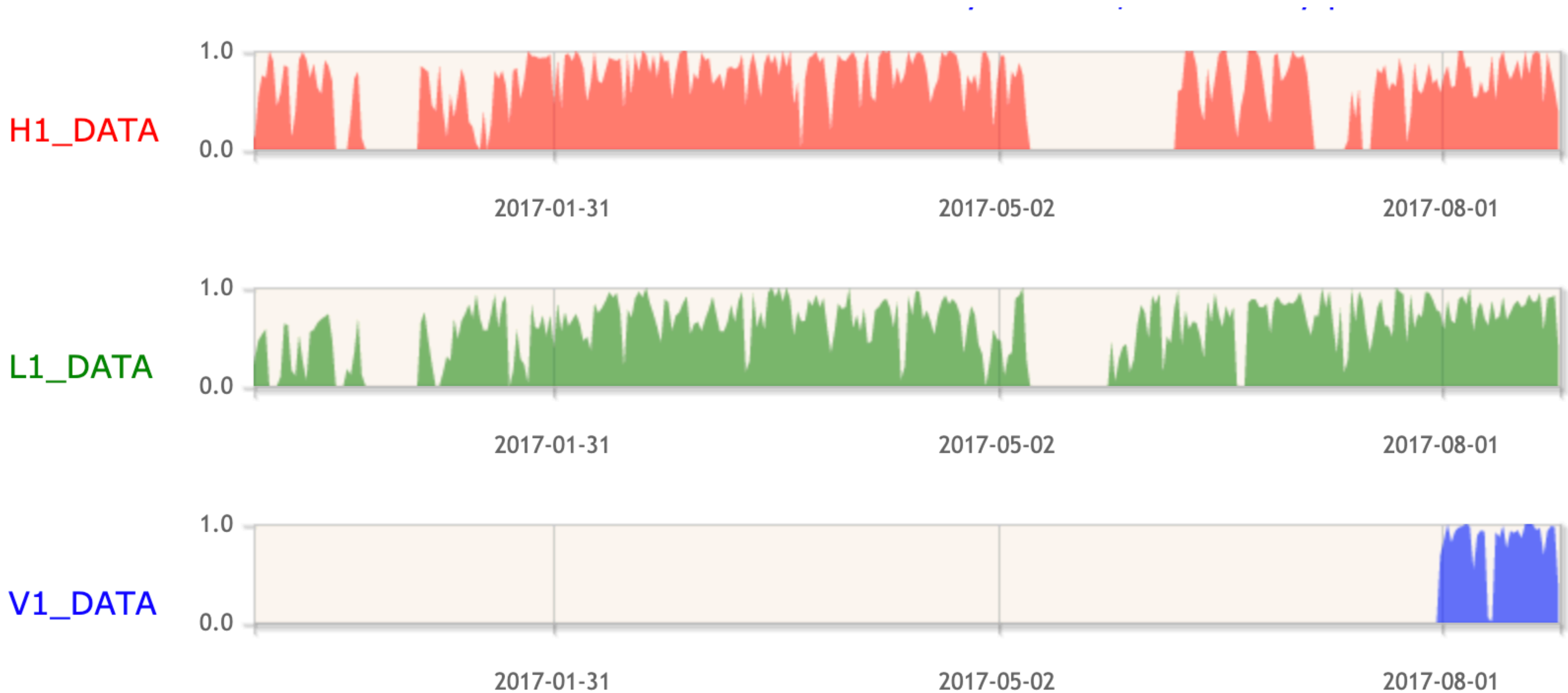


How to get Data Quality Segments

Using GWOSC Timeline Query

https://www.gw-openscience.org/timeline/query/Run/

Here I selected the O2 data flags for H1, L1, and V1



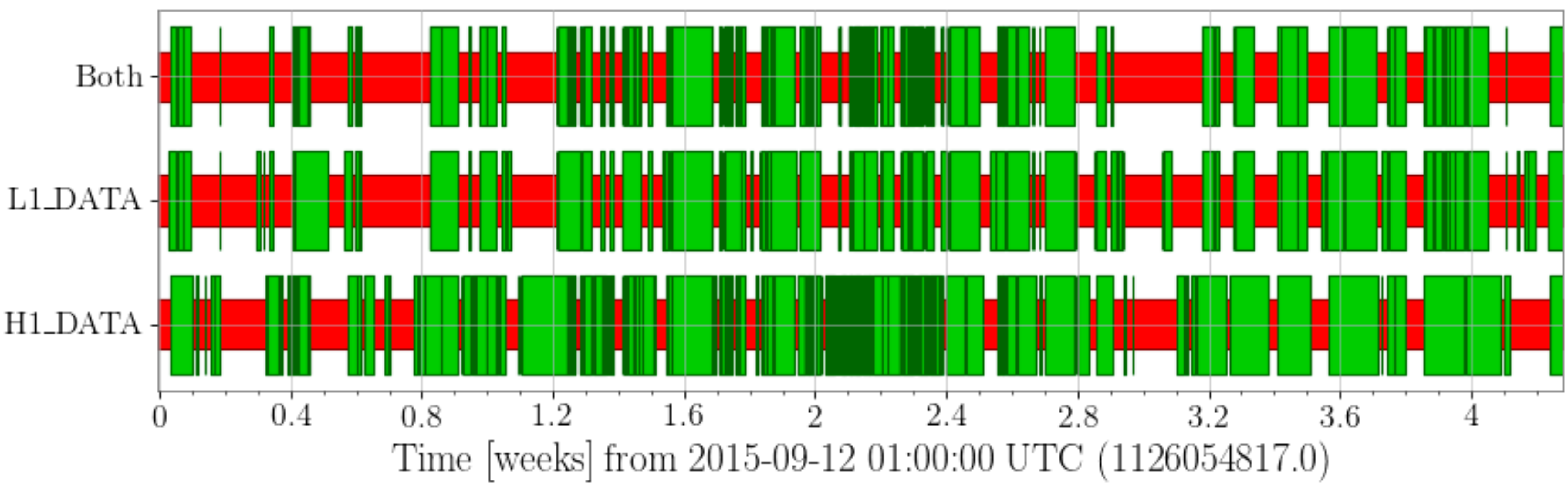
Segments can be plotted (with interactive zooming) or downloaded

How to get Data Quality Segments

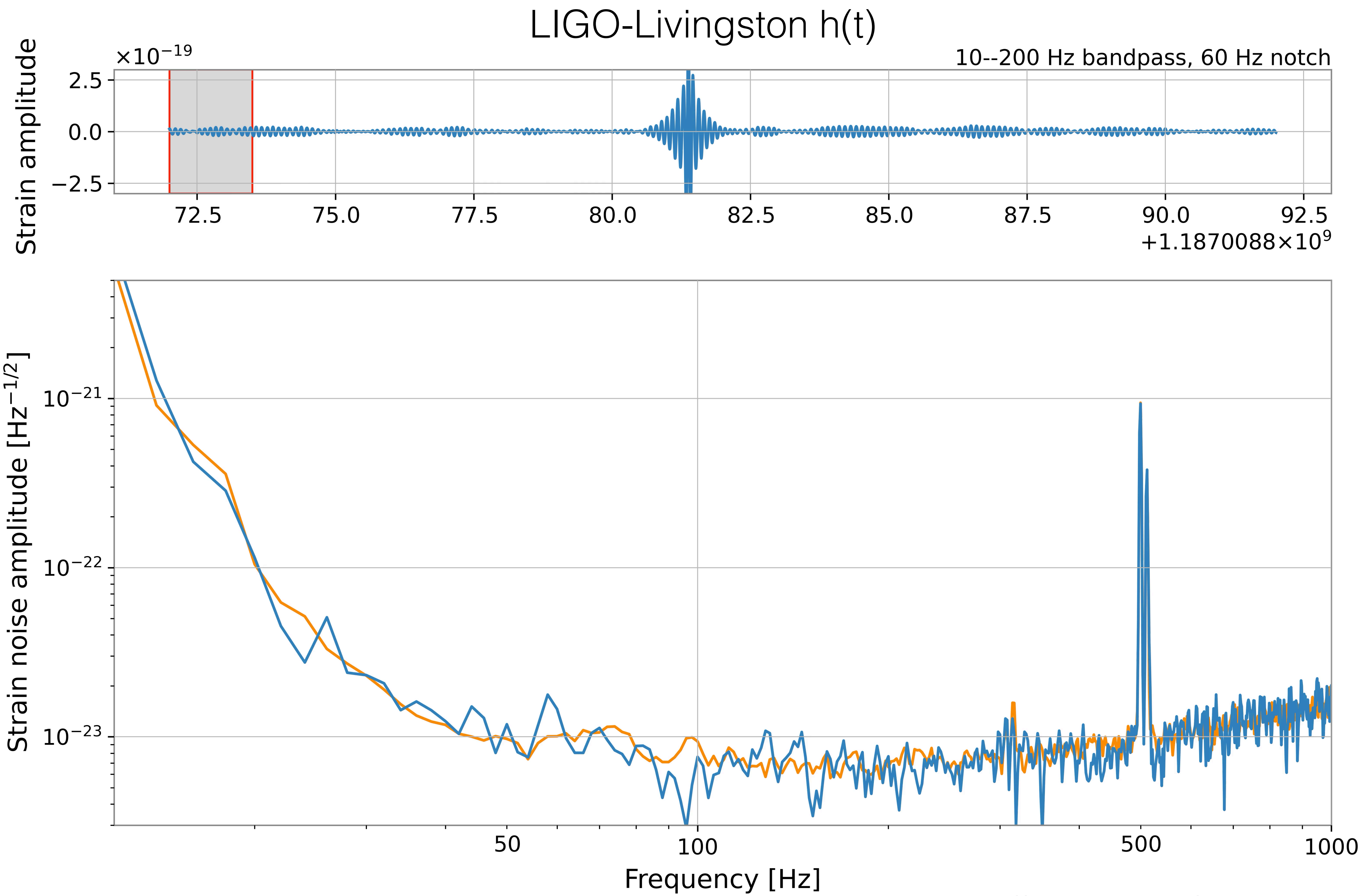
Using GWpy

Example showing how to find and plot data quality segments from O1:

https://gwpy.github.io/docs/stable/examples/segments/open-data.html



Search-specific data quality segments
Example: GW170817

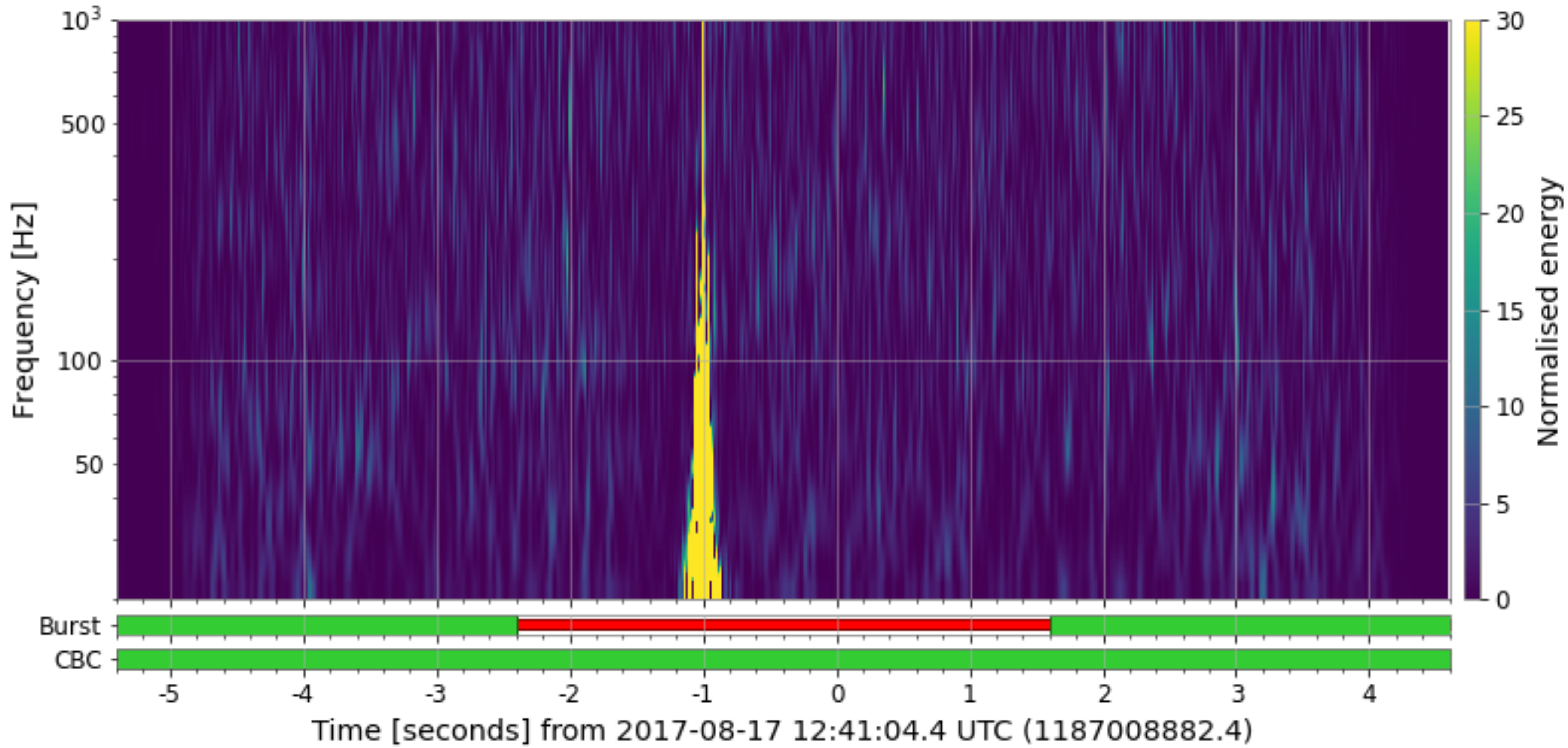


Modified from Duncan's original script. Notebook: https://github.com/marissawalker/data-quality-odw-2020/blob/master/LIGO_data_animation.ipynb

0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

Search-specific data quality segments
Example: GW170817

LIGO-Livingston h(t)



Notebook: https://github.com/marissawalker/data-quality-odw-2020/blob/master/GW170817_glitch_qscans.ipynb

In Tutorial 1.3 you will learn to make a version of this plot that will show the BNS signal in this plot, not just the glitch

Data quality for event validation: examples from GWTC-1

GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs

B.P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration)

Phys. Rev. X **9**, 031040 – Published 4 September 2019

DOI: https://doi.org/10.1103/PhysRevX.9.031040

See talks by Derek Davis and Alan Weinstein tomorrow for details on the searches for compact binary coalescences (CBC)

Example: MC170720

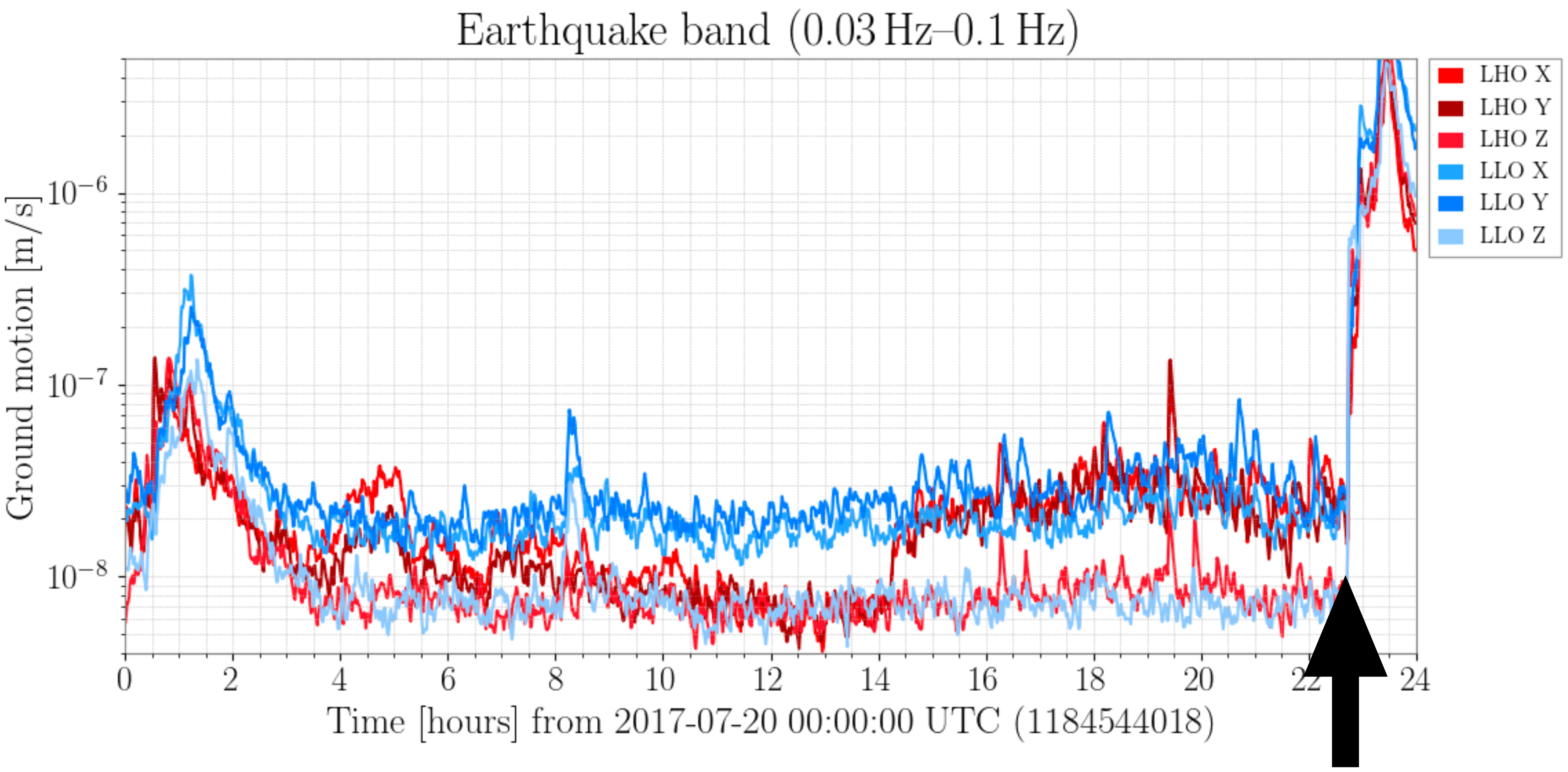
Marginal trigger most likely of terrestrial origin

Within 20 s of trigger 170720, excess ground motion from earthquakes forced the Livingston detector to drop out of its nominal mode of operation. Before the detector dropped out of the observing state, the data are heavily polluted with scattering artifacts that could account for the SNR of the triggers. As the PyCBC search does not consider times near the edges of observing periods, this time period is also not analyzed by that search. Artifacts related to scattered light are also observed at Hanford at this time.

GWTC-1 p. 11

Example notebook: https://github.com/marissawalker/data-quality-odw-2020/blob/master/MC170720.ipynb

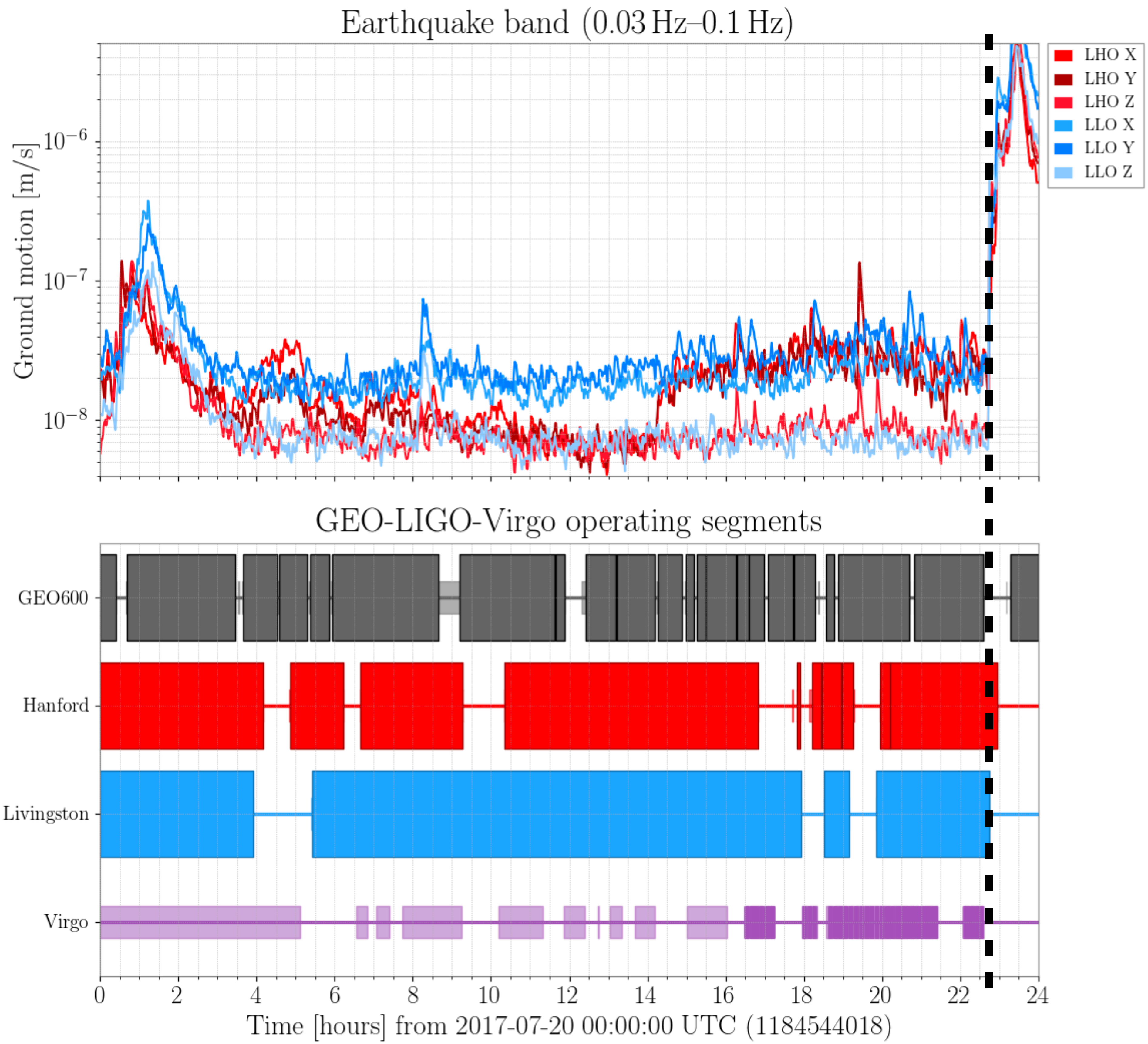
Example: MC170720



https://www.gw-openscience.org/detector_status/day/20170720/environment/ground_motion/

Within 20 s of trigger 170720, excess ground motion from earthquakes forced the Livingston detector to drop out of its

Example: MC170720



Example: MC170720

Marginal trigger most likely of terrestrial origin

marginal trigger. Because of the short observing duration, this time period is not analyzed by the PyCBC search.

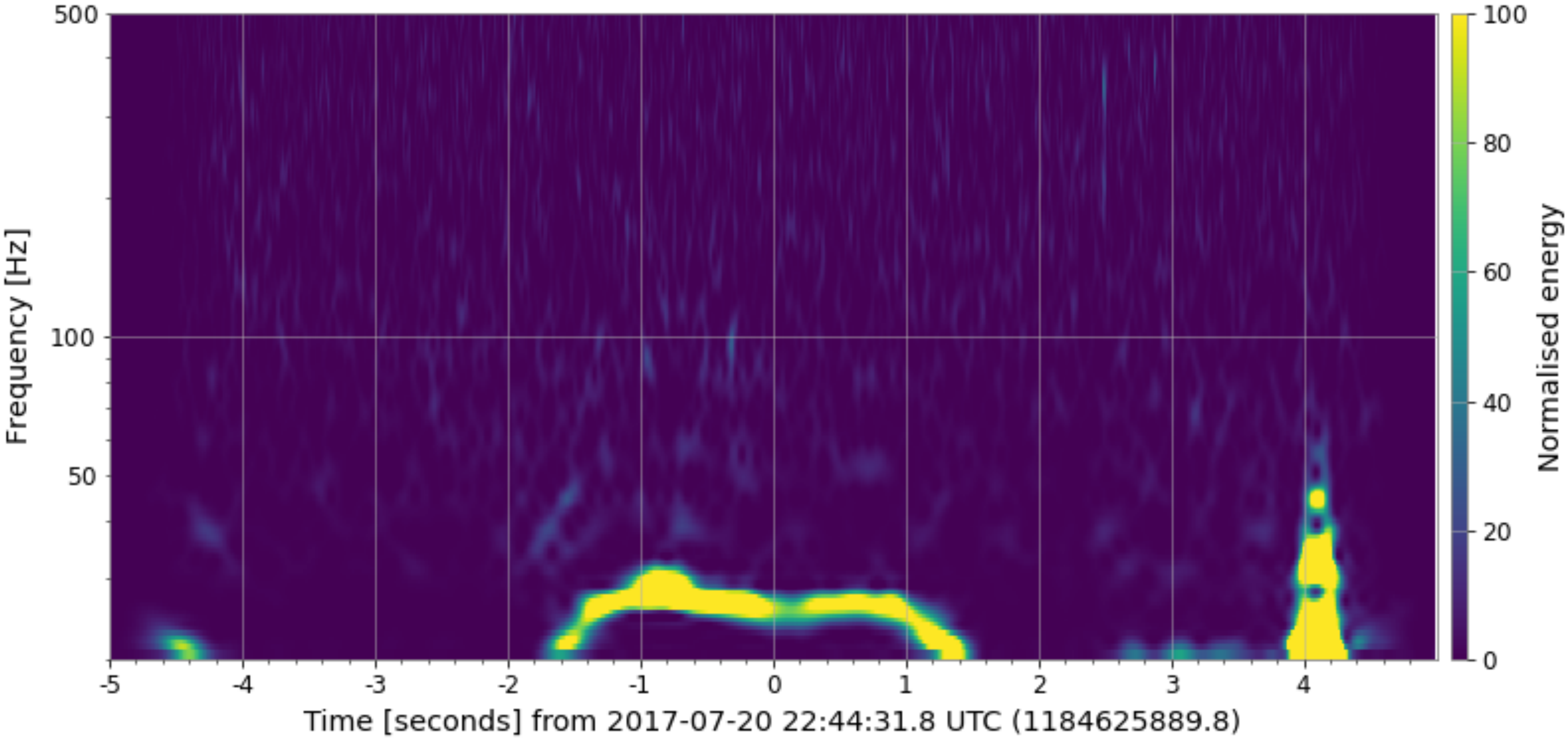
Within 20 s of trigger 170720, excess ground motion from earthquakes forced the Livingston detector to drop out of its nominal mode of operation. Before the detector dropped out of the observing state, the data are heavily polluted with scattering artifacts that could account for the SNR of the triggers. As the PyCBC search does not consider times near the edges of observing periods, this time period is also not analyzed by that search. Artifacts related to scattered light are also observed at Hanford at this time.

GWTC-1 p. 11

Example: MC170720

Marginal trigger most likely of terrestrial origin

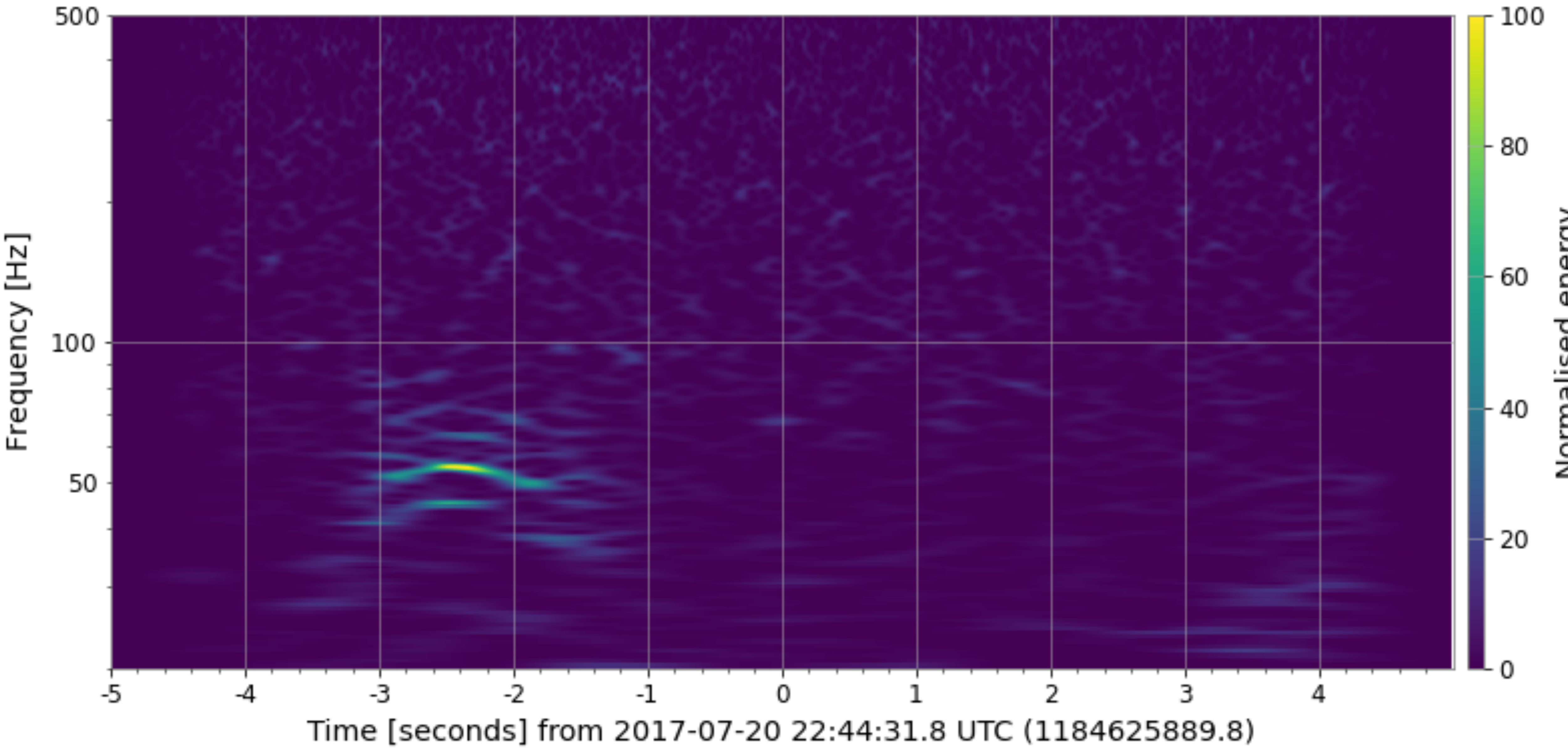
Livingston



Example: MC170720

Marginal trigger most likely of terrestrial origin

Hanford



Useful data quality references

For glitches:
GW150914 Detector characterization paper: arXiv 1602.03844
O1 CBC DQ paper; CQG (2018): http://iopscience.iop.org/article/10.1088/1361-6382/aaaafa/meta
Gravity Spy: gravityspy.org

For lines:
O1/O2 lines paper: arXiv 1801.07204
O2 lines catalog on the GWOSC: https://www.gw-openscience.org/o2speclines/

Data quality segments:
Data quality timelines: https://www.gw-openscience.org/timeline/

Public interferometer status monitoring: https://www.gw-openscience.org/detector_status/

O3 public alerts: https://gracedb.ligo.org/superevents/public/O3/

Example notebooks used in this talk:
https://github.com/marissawalker/data-quality-odw-2020

GWpy documentation: https://gwpy.github.io/

Thank you!

Example: MC161217

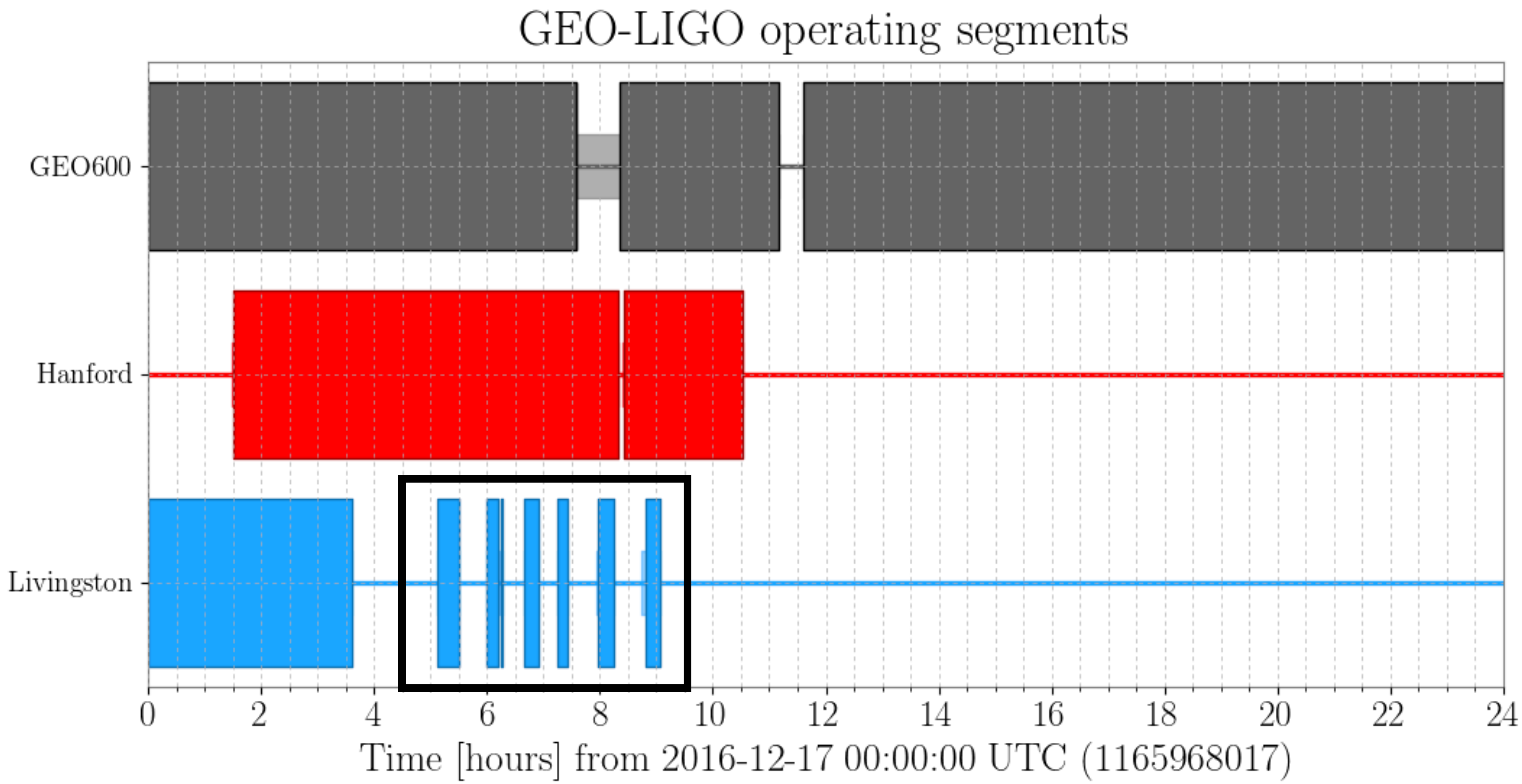
Marginal trigger most likely of terrestrial origin

the observatories at the times in question identified high-amplitude ground motion correlated with the scattering.

The marginal trigger 161217 occurs during a period of high-amplitude ground motion at Livingston caused by storm activity. During this storm activity, the Livingston detector is not able to maintain a stable interferometer for periods longer than 10 min. The presence of intense scattering artifacts contribute to the unstable state of the interferometer and could account for the SNR of the marginal trigger. Because of the short observing duration, this time period is not analyzed by the PyCBC search.

GWTC-1 p. 11

Example: MC161217



https://www.gw-openscience.org/detector_status/day/20161217/instrument_performance/analysis_time/

storm activity. During this storm activity, the Livingston detector is not able to maintain a stable interferometer for periods longer than 10 min. The presence of intense

Example: MC161217

Marginal trigger most likely of terrestrial origin

the observatories at the times in question identified high-amplitude ground motion correlated with the scattering.

The marginal trigger 161217 occurs during a period of high-amplitude ground motion at Livingston caused by storm activity. During this storm activity, the Livingston detector is not able to maintain a stable interferometer for periods longer than 10 min. The presence of intense scattering artifacts contribute to the unstable state of the interferometer and could account for the SNR of the marginal trigger. Because of the short observing duration, this time period is not analyzed by the PyCBC search.

Within 20 s of trigger 170720, excess ground motion from

GWTC-1 p. 11

Try it yourself! MC161217 data quality

Using the example notebook about MC170720, explore the data quality of MC161217, and compare with the statements made in GWTC-1:

- Check the length of Livingston’s data segment
- Using the q transform, make a spectrogram to look for evidence of scattering