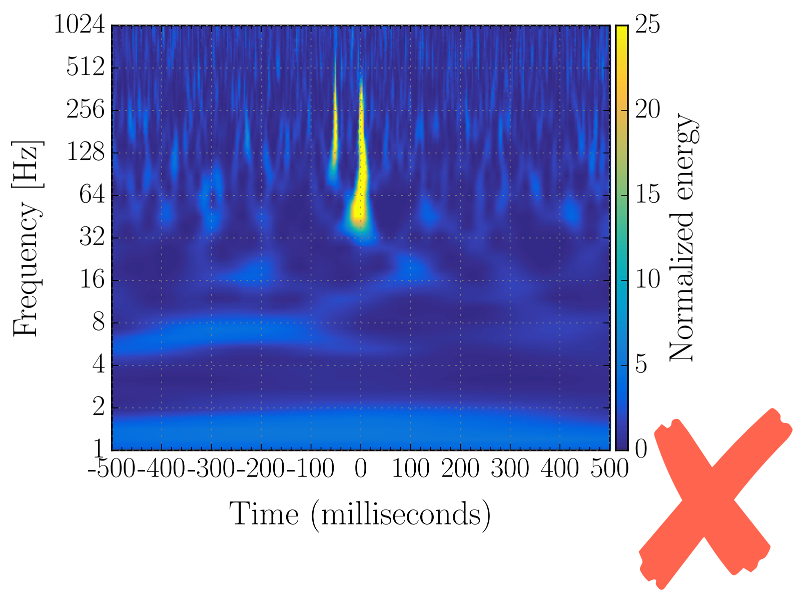
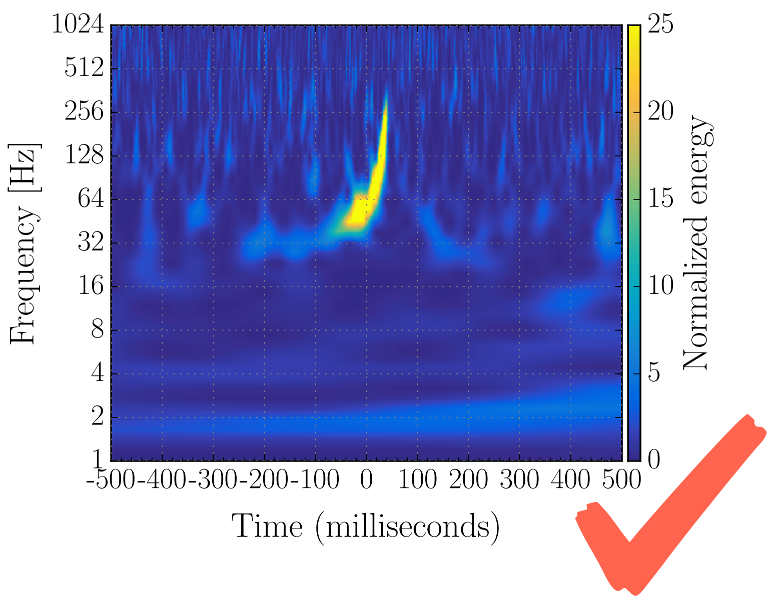
Data Quality and GWpy





Laura Nuttall

Gravitational Wave Open Data Workshop #4

Outline

- What is GWpy?
- What does GW data look like?
  - Time domain
  - Frequency domain
  - Time-frequency representation
  - Noise Subtraction
- Data Quality
  - Various forms of noise
- Identifying and Mitigating Sources of Noise
  - Data Quality Vetoes
  - Event Validation
- Data Quality Information
  - Detector Status Summary Pages
- Summary of resources and references

A big thank you to previous workshop presenters D Macleod, J McIver and M Walker. I have taken a lot of inspiration for these slides from your great work!

Fork me on GitHub

GWpy

a python package for gravitational-wave astrophysics.

>>> help(gwpy)

Project status

pypi package 2.0.4 | build passing | coverage 91%

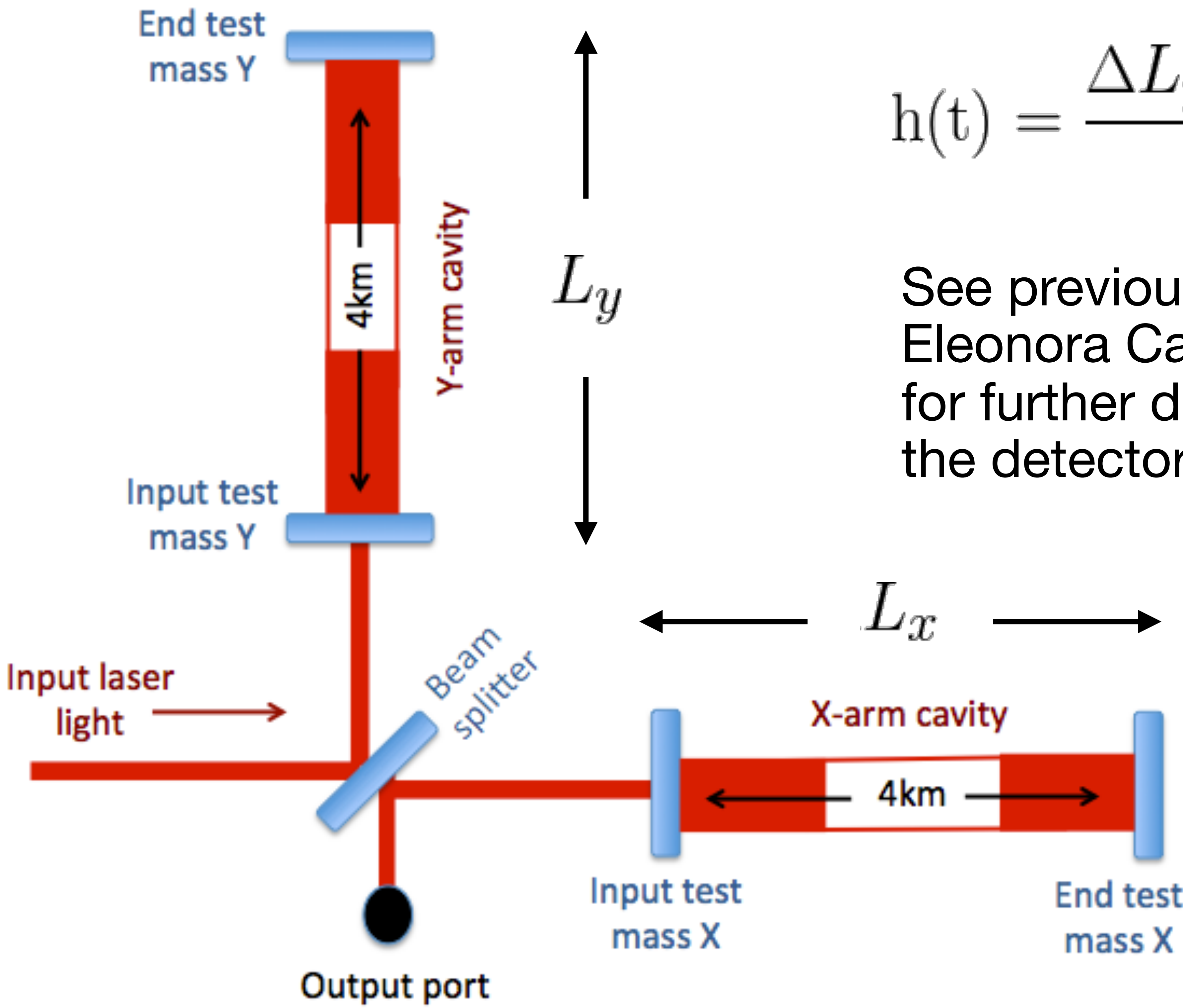
Useful links

Installation | Report an issue

The GWpy package contains classes and utilities providing tools and methods for studying data from gravitational-wave detectors, for astrophysical or instrumental purposes.

Documentation at: https://gwpy.github.io/docs/stable/index.html

What is strain data, h(t) ?

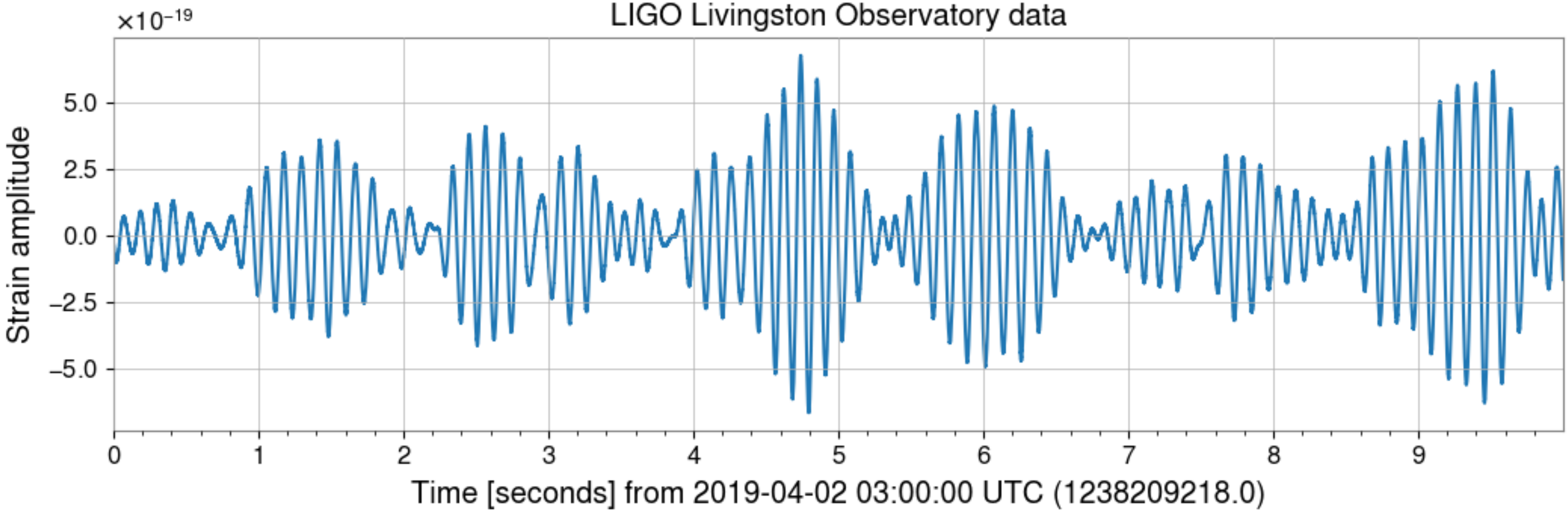


$$\mathrm{h(t)} = \frac{\Delta L_y - \Delta L_x}{L}$$

See previous talk by Eleonora Capocasa for further details on the detectors!

Slide by J McIver

What does GW data look like?

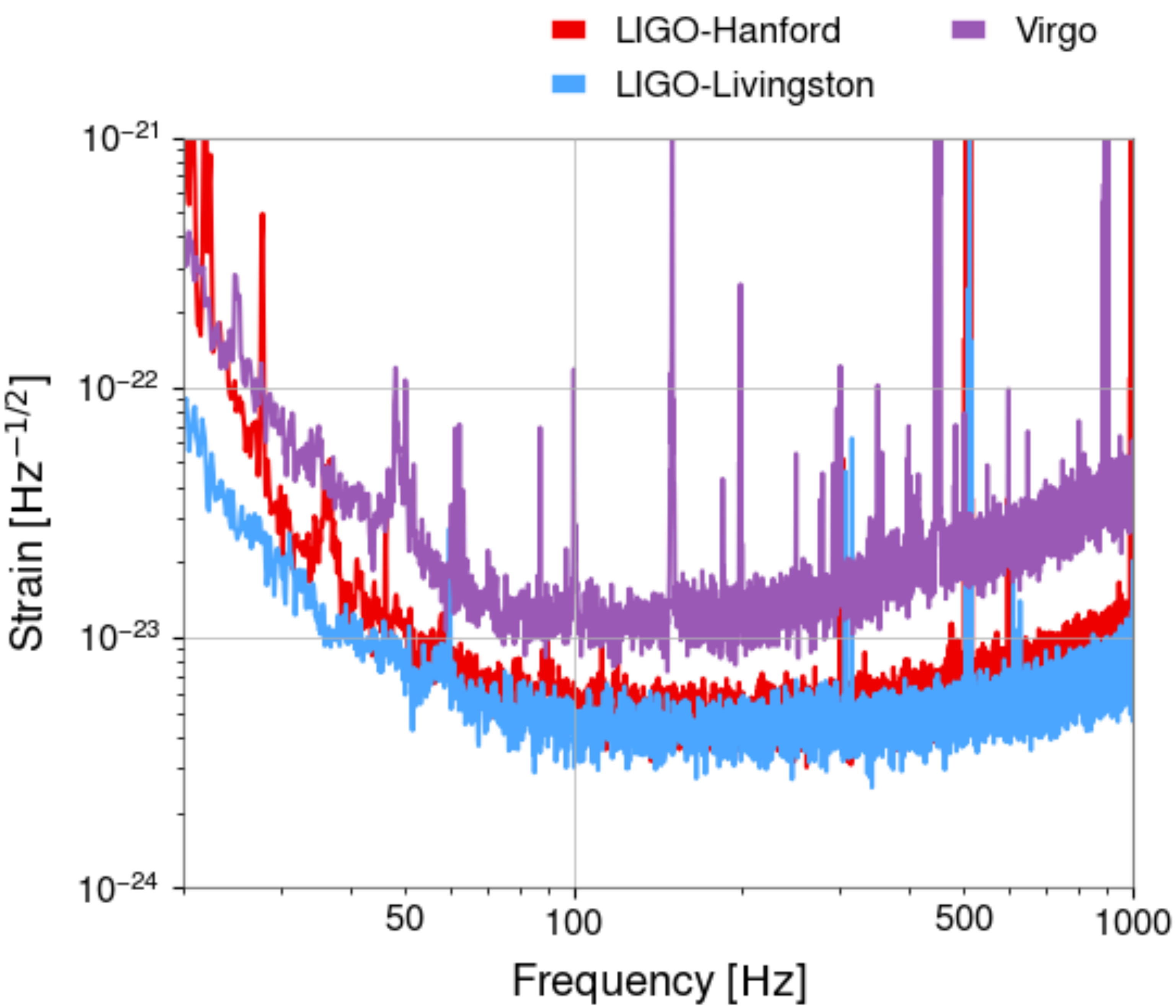


Open data: LIGO and Virgo data are sampled at either 16384 or 4096 Hz

Users should choose the sampling rate which is most appropriate for your search

In tutorials 1.1 and 1.2 you will learn how to get and plot the data

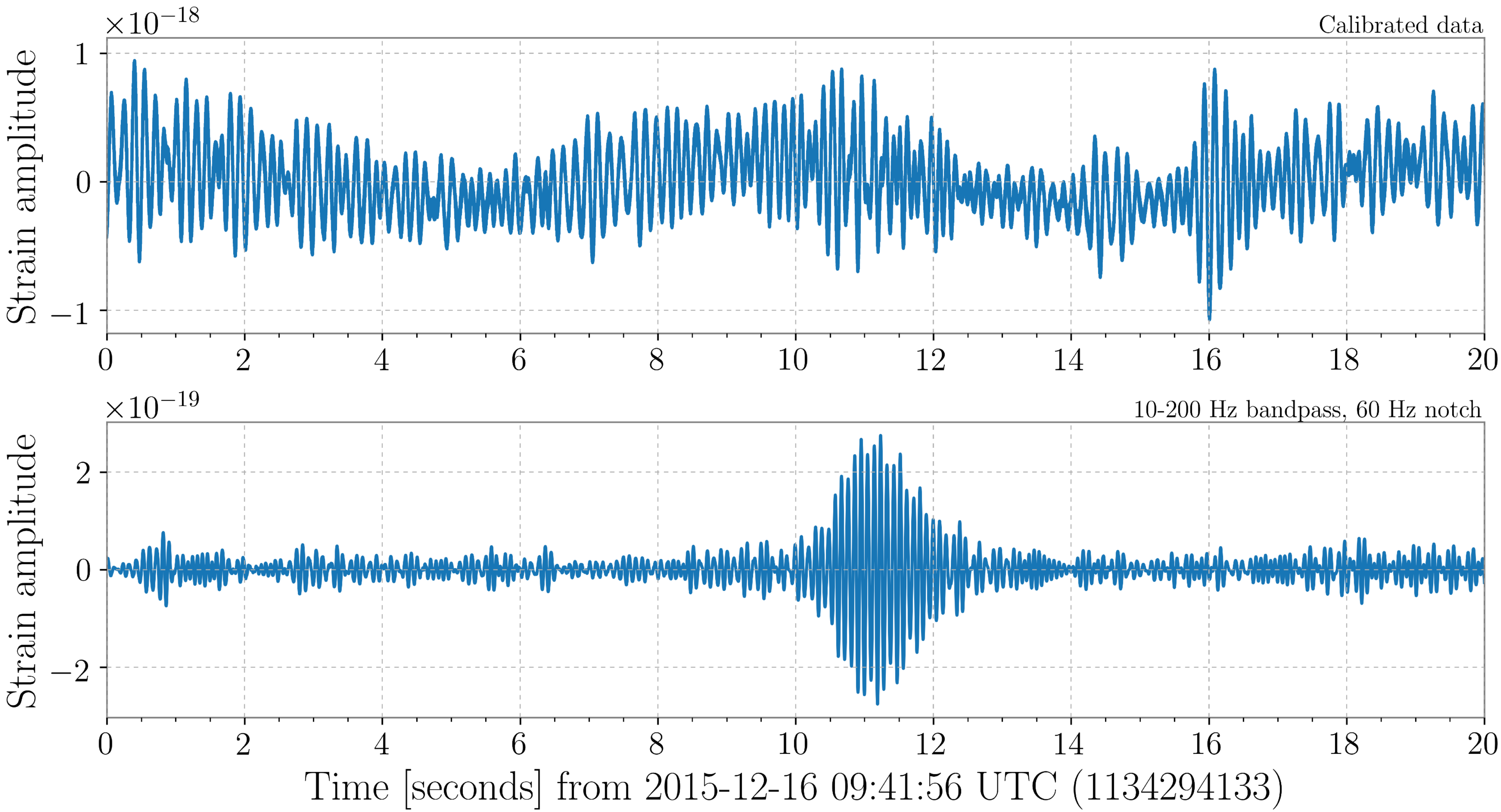
Data in the Frequency domain



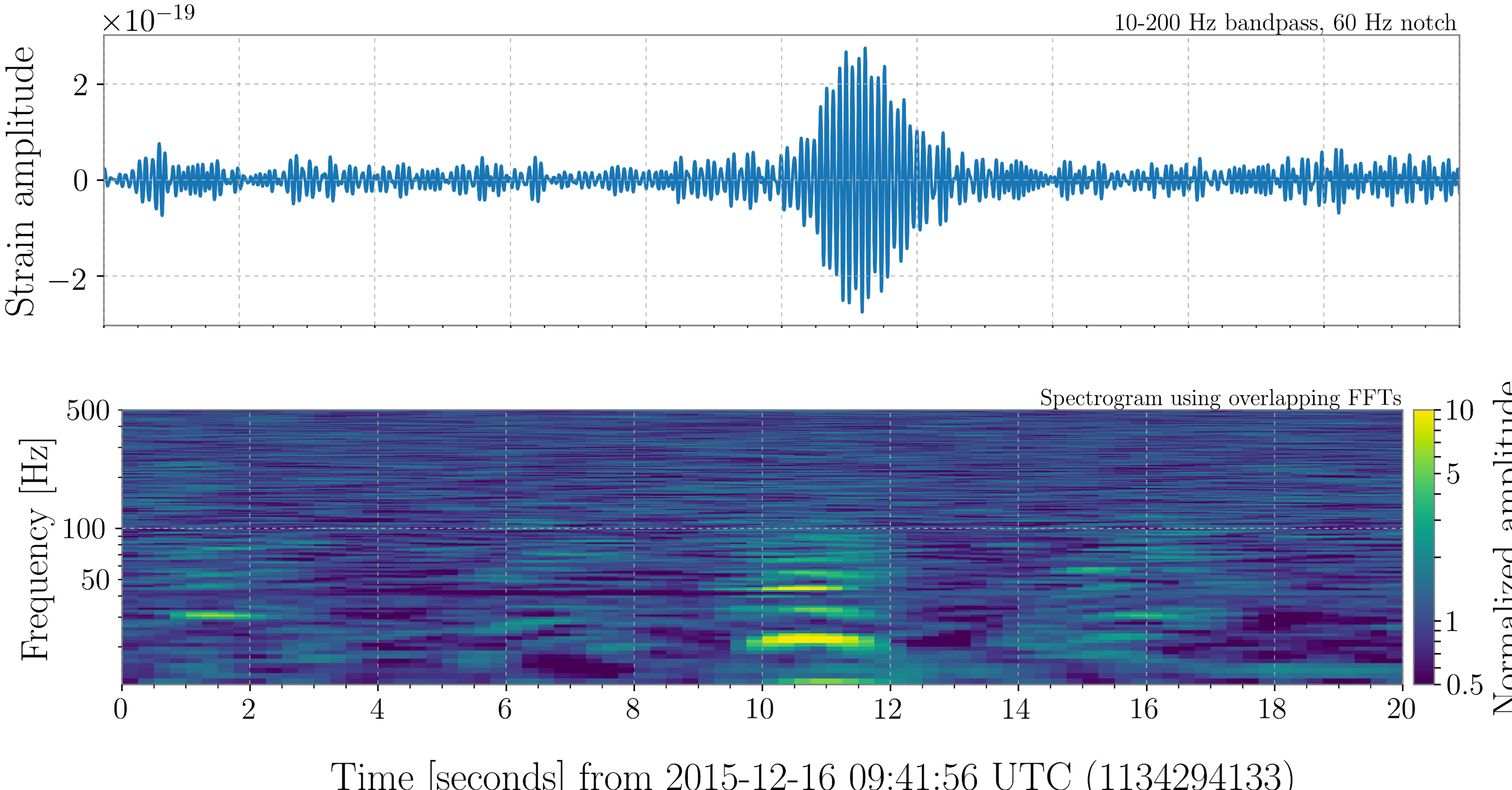
Amplitude Spectral Density (ASD) of data from all three detectors on 2nd April 2019 at 03:00 UTC

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

Features in the GW data

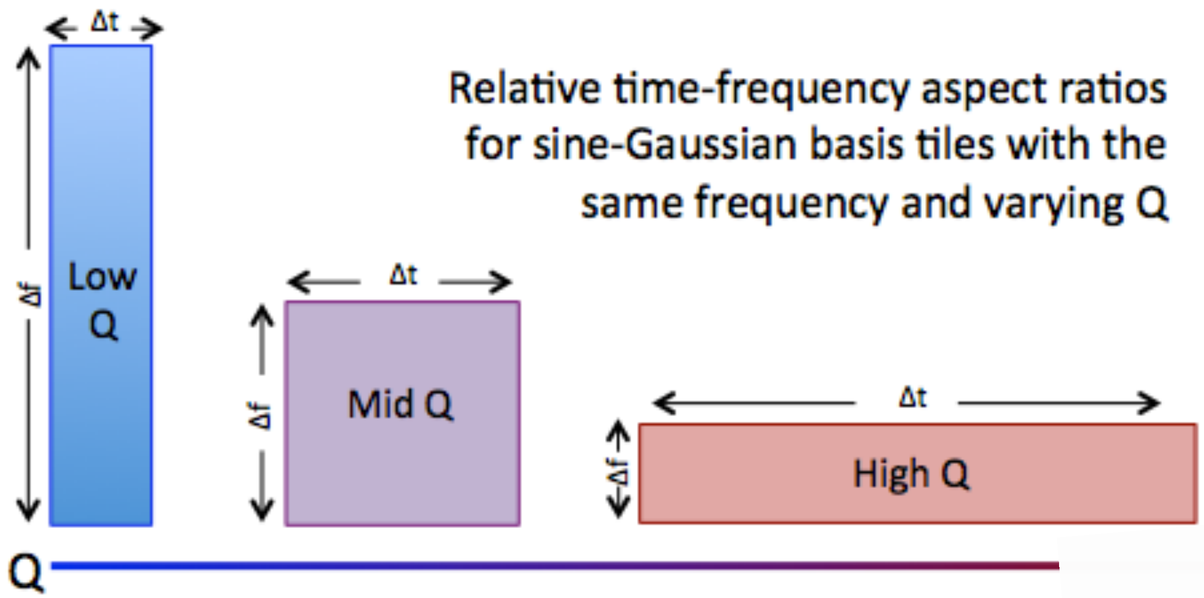


Features in the GW data: Spectrograms

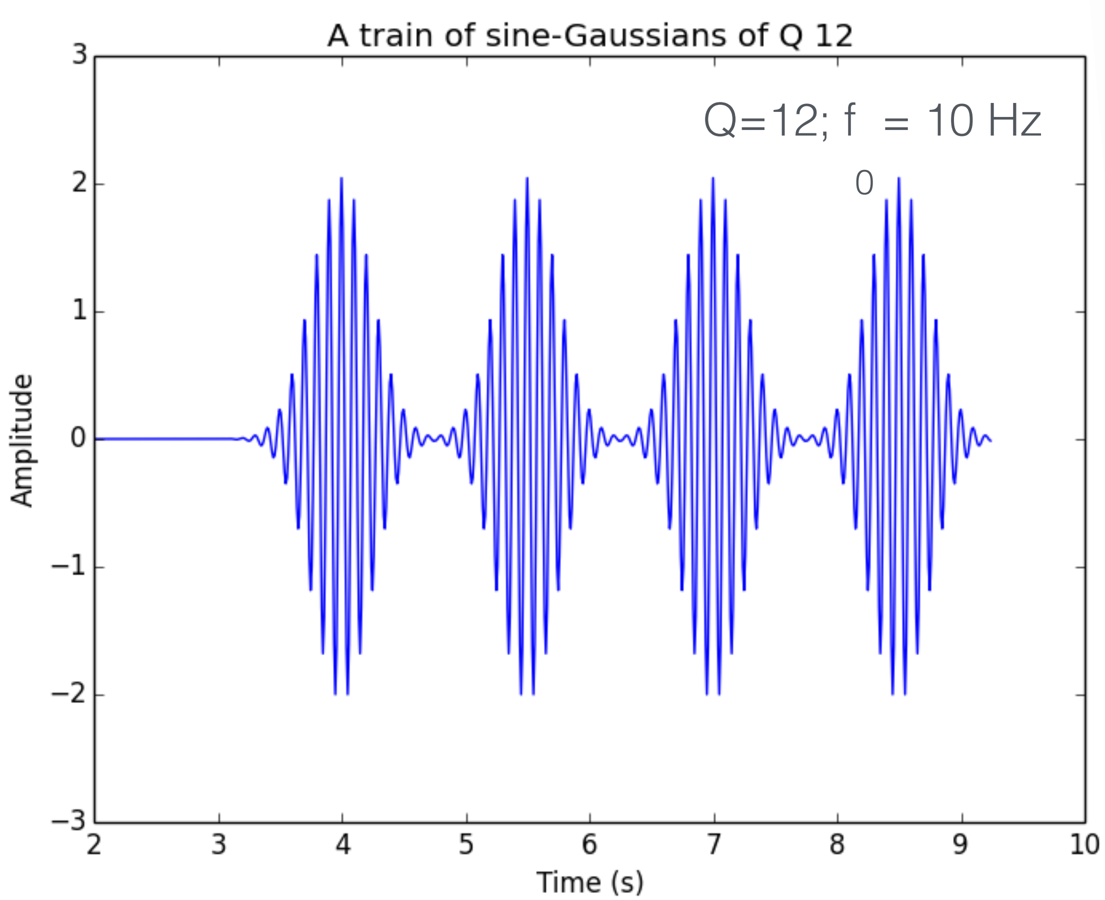


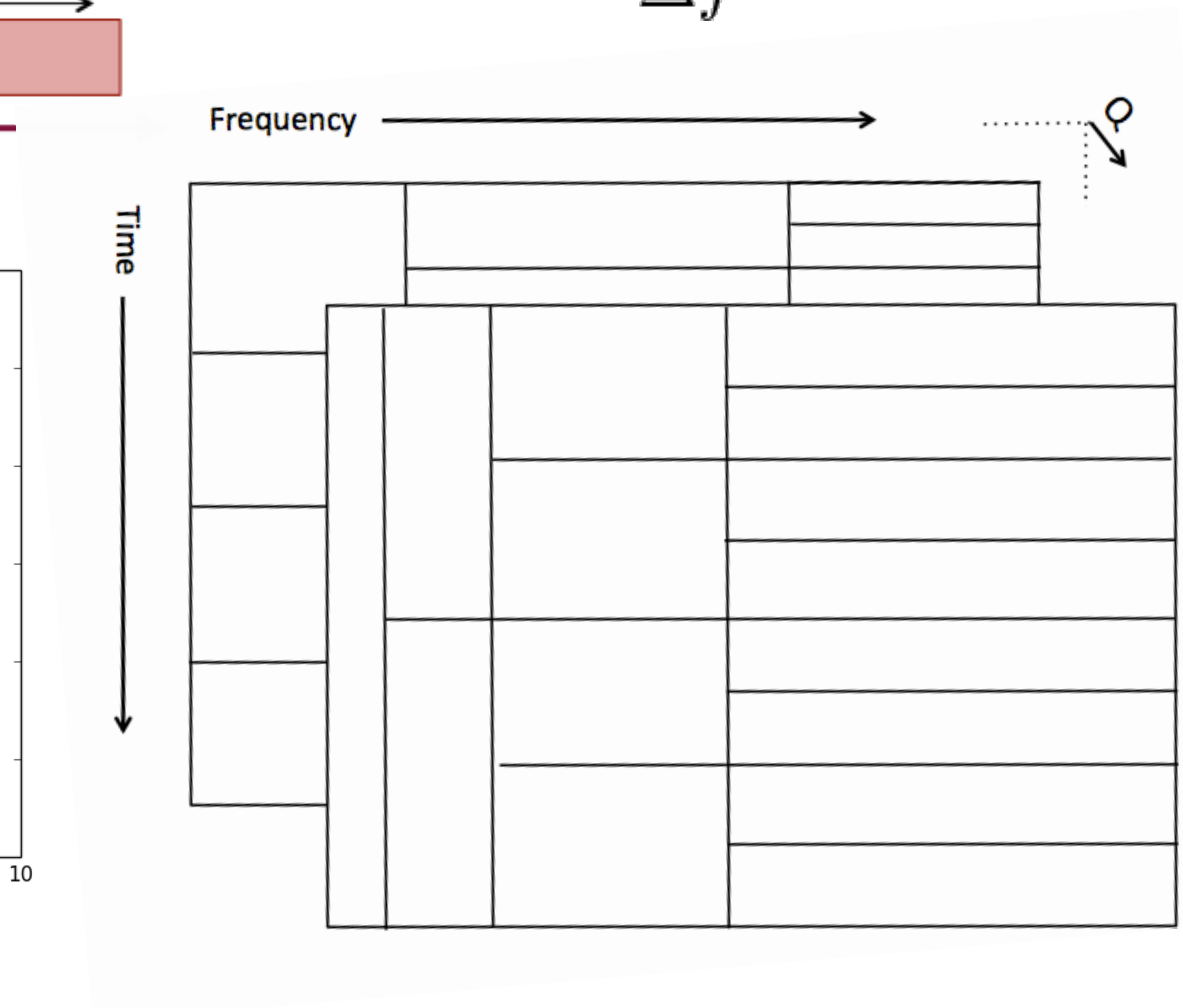
The Q transform

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



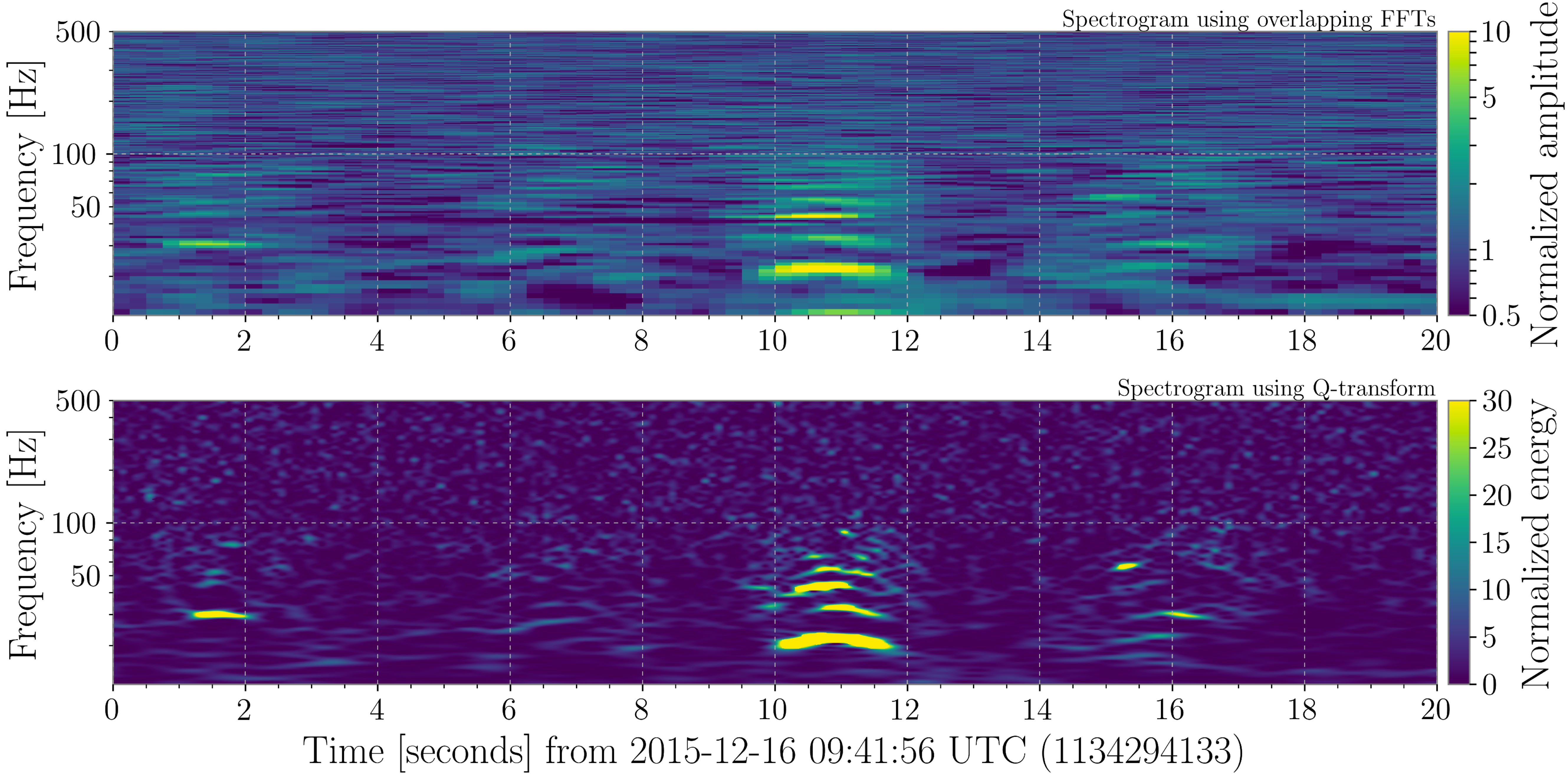
$$Q = \frac{f_0}{\Delta f}$$





S. Chatterji et al., *Multiresolution techniques for the detection of gravitational-wave bursts*, CQG **21** S1809 (2004)

Features in the GW data: Q transform



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

Noise Subtraction

After data collection we remove several independently measured terrestrial contributions to the detector noise:

- LIGO - remove calibration lines and 60Hz AC power mains harmonics. We also remove some additional noise due to non stationary couplings

- Virgo - remove broadband noise, including frequency noise from the laser, noise introduced when controlling the displacement of the beam splitter and amplitude noise of the 56 MHz modulation frequency.

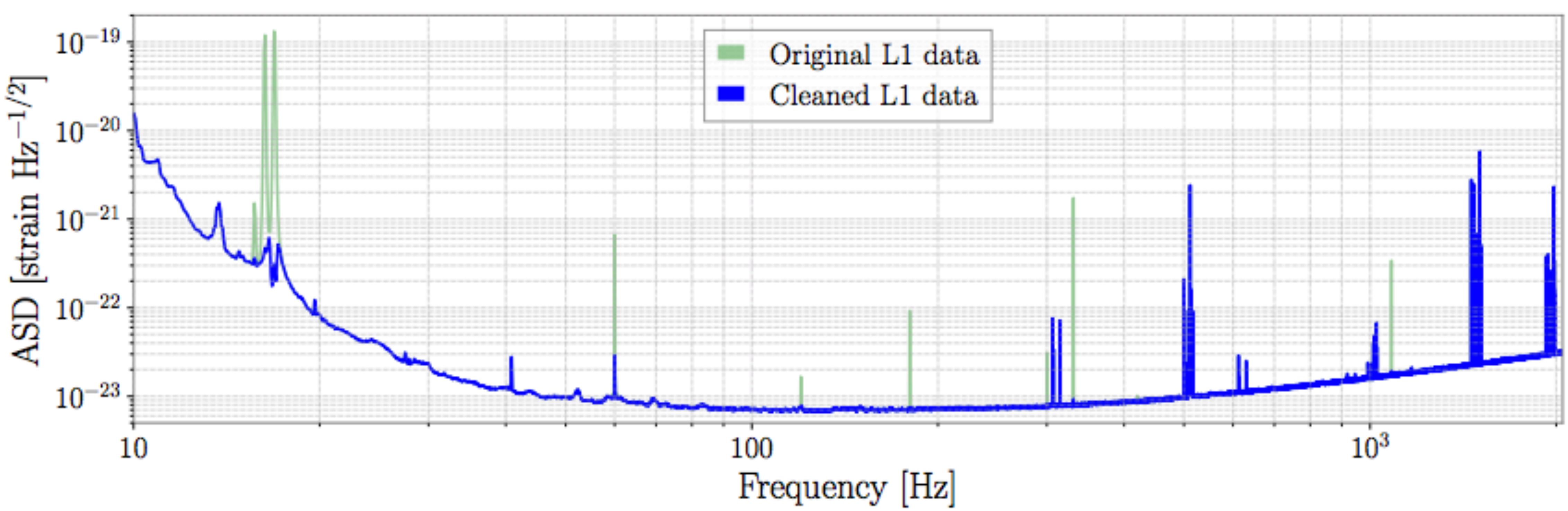


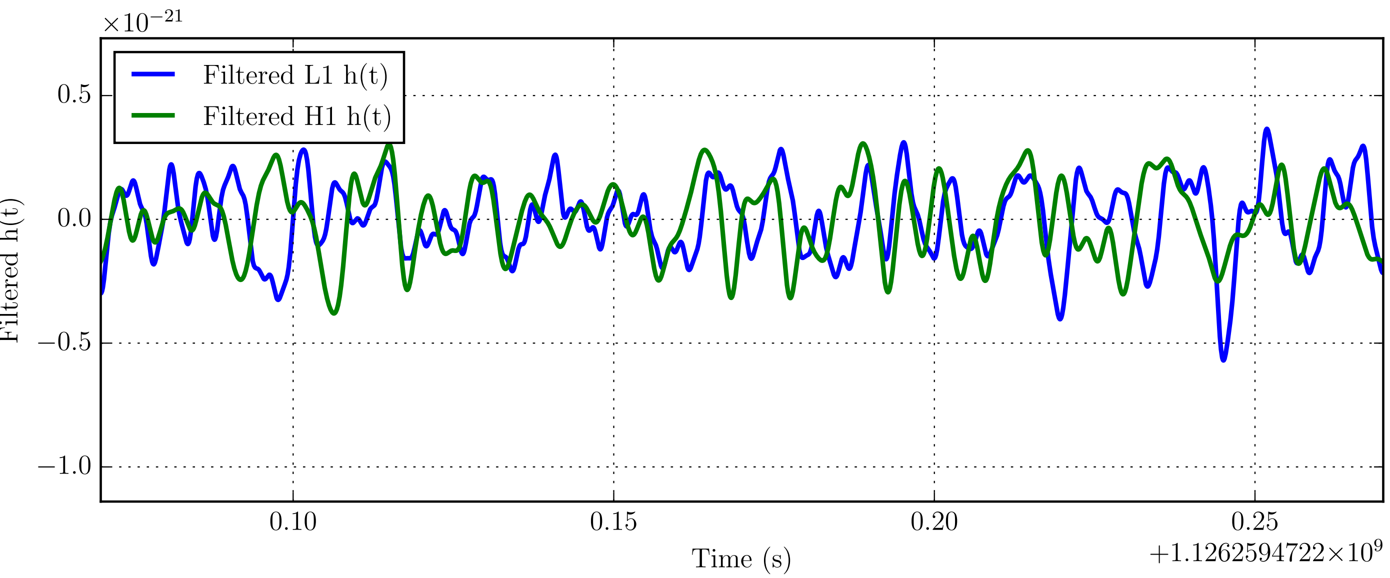
Image: D. Davis et al., *Improving the sensitivity of Advanced LIGO using noise subtraction*, CQG **36** 055011 (2019)

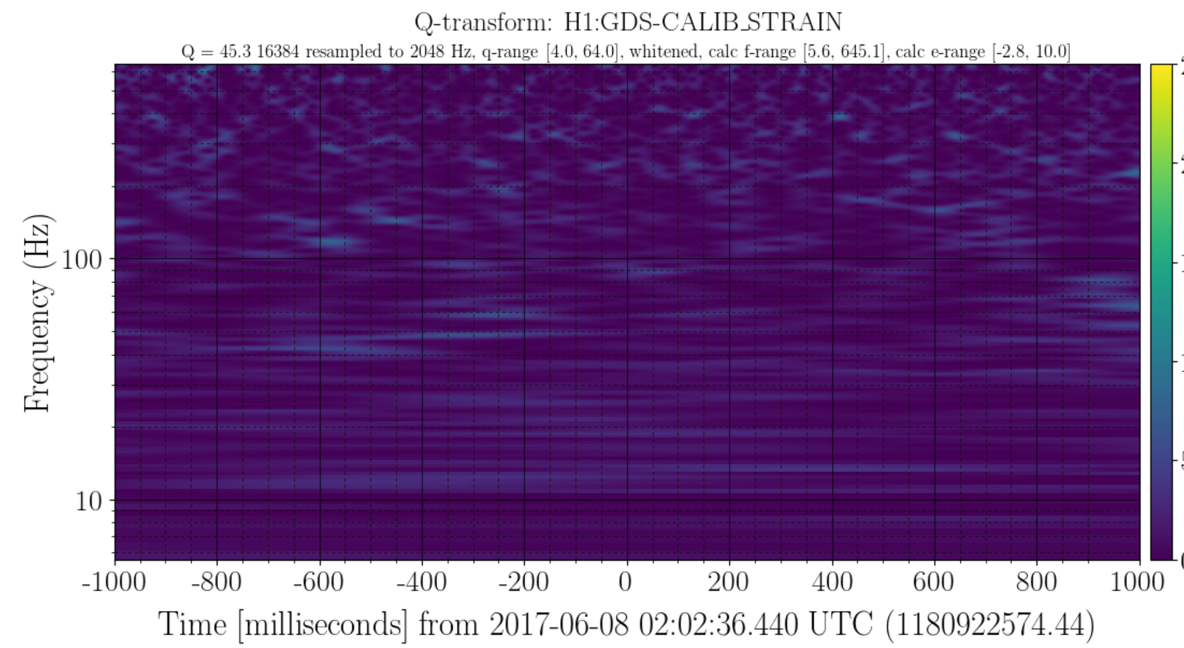
Outline

- What is GWpy?
- What does GW data look like?
  - Time domain
  - Frequency domain
  - Time-frequency representation
  - Noise Subtraction
- Data Quality
  - Various forms of noise
- Identifying and Mitigating Sources of Noise
  - Data Quality Vetoes
  - Event Validation
- Data Quality Information
  - Detector Status Summary Pages
- Summary of resources and references

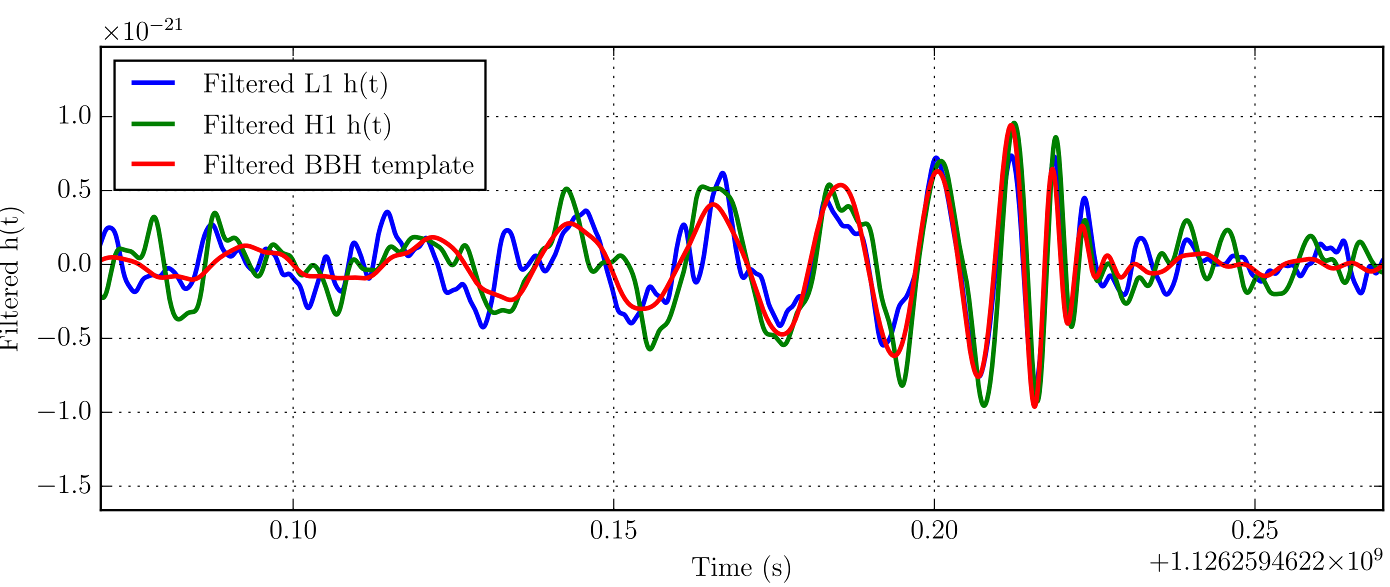
GW data in a perfect world…

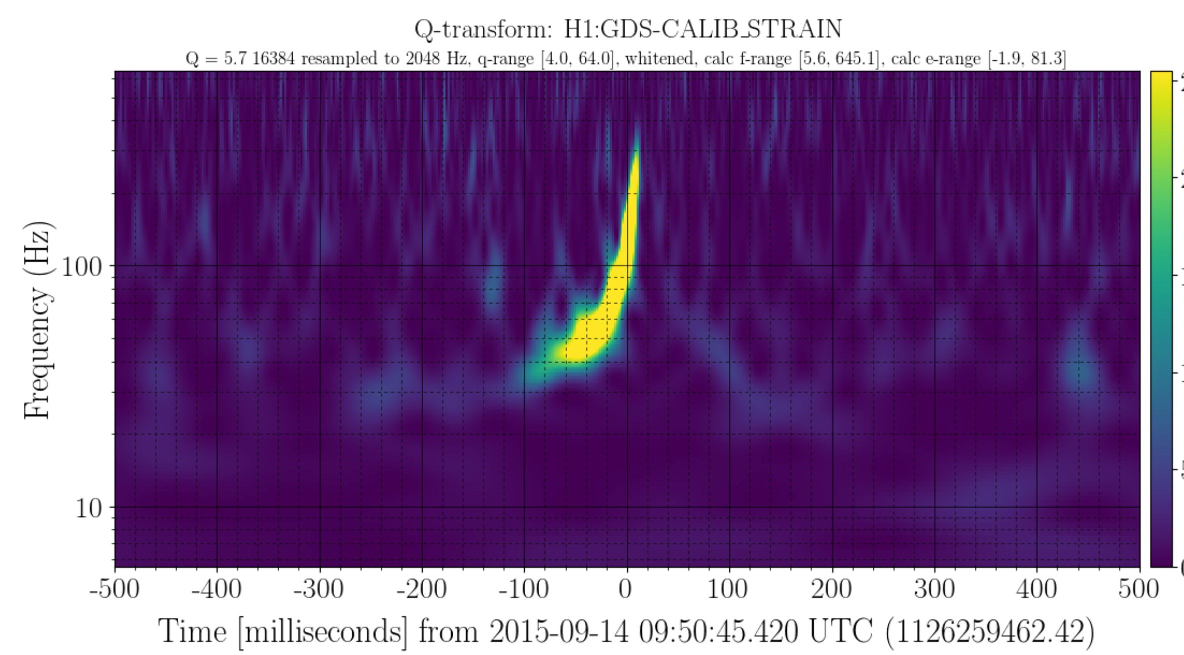
No signal





Signal

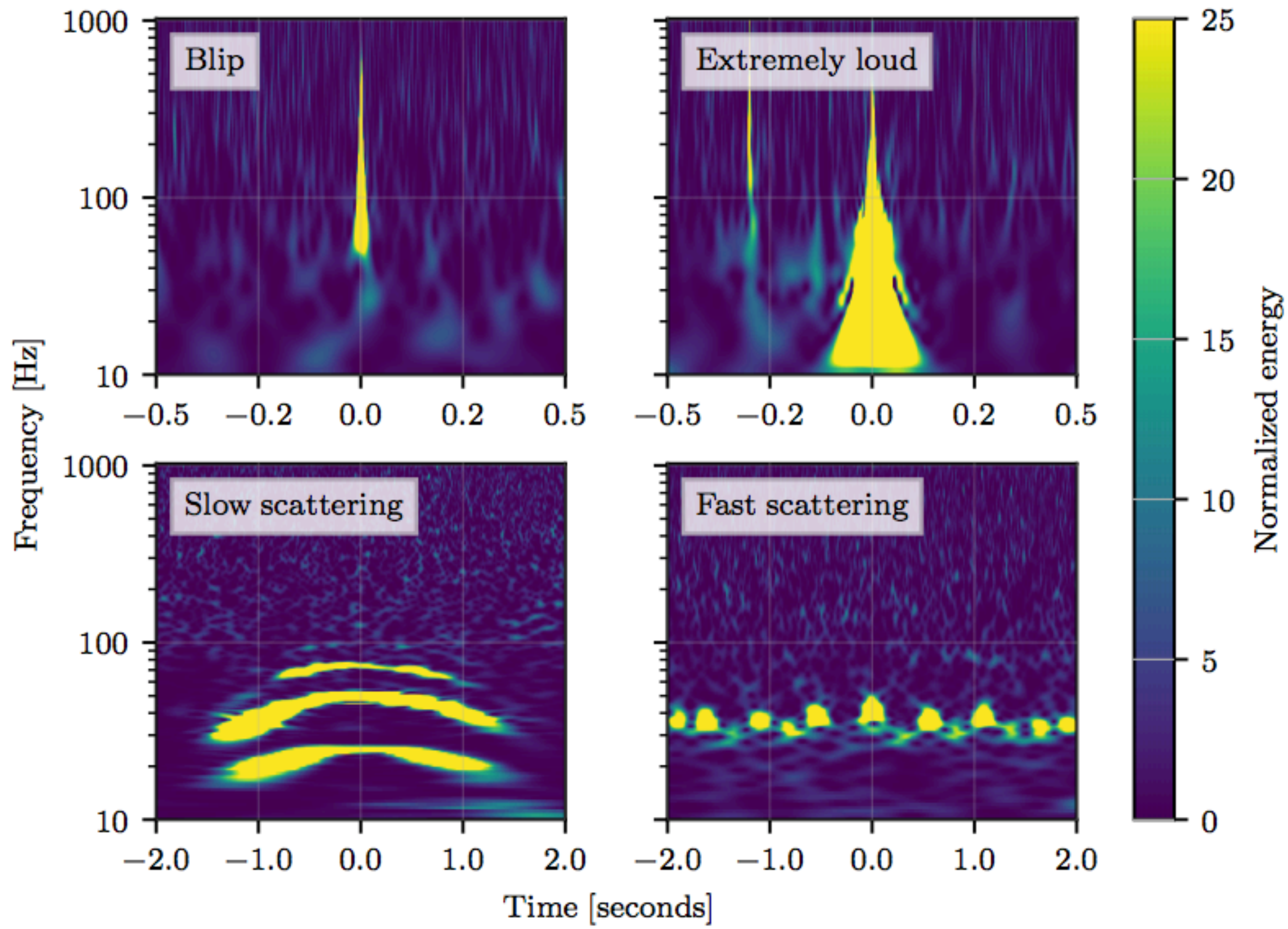




Slide by M Walker

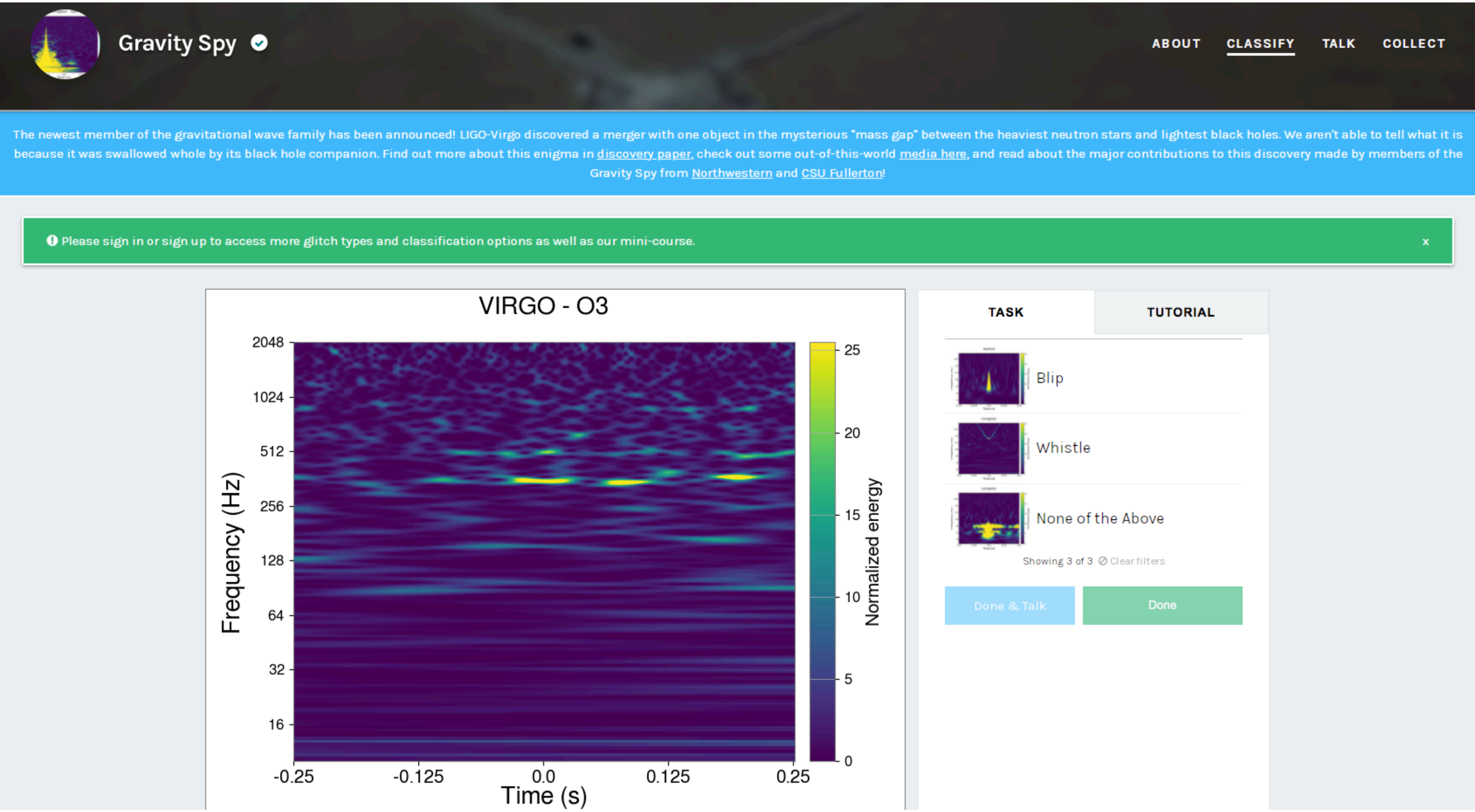
In reality…GW data also contains instrumental and environmental artefacts

Glitches are non-gaussian noise transients that are found in all detectors. Below are some of the worst glitch classes in LIGO data from O3



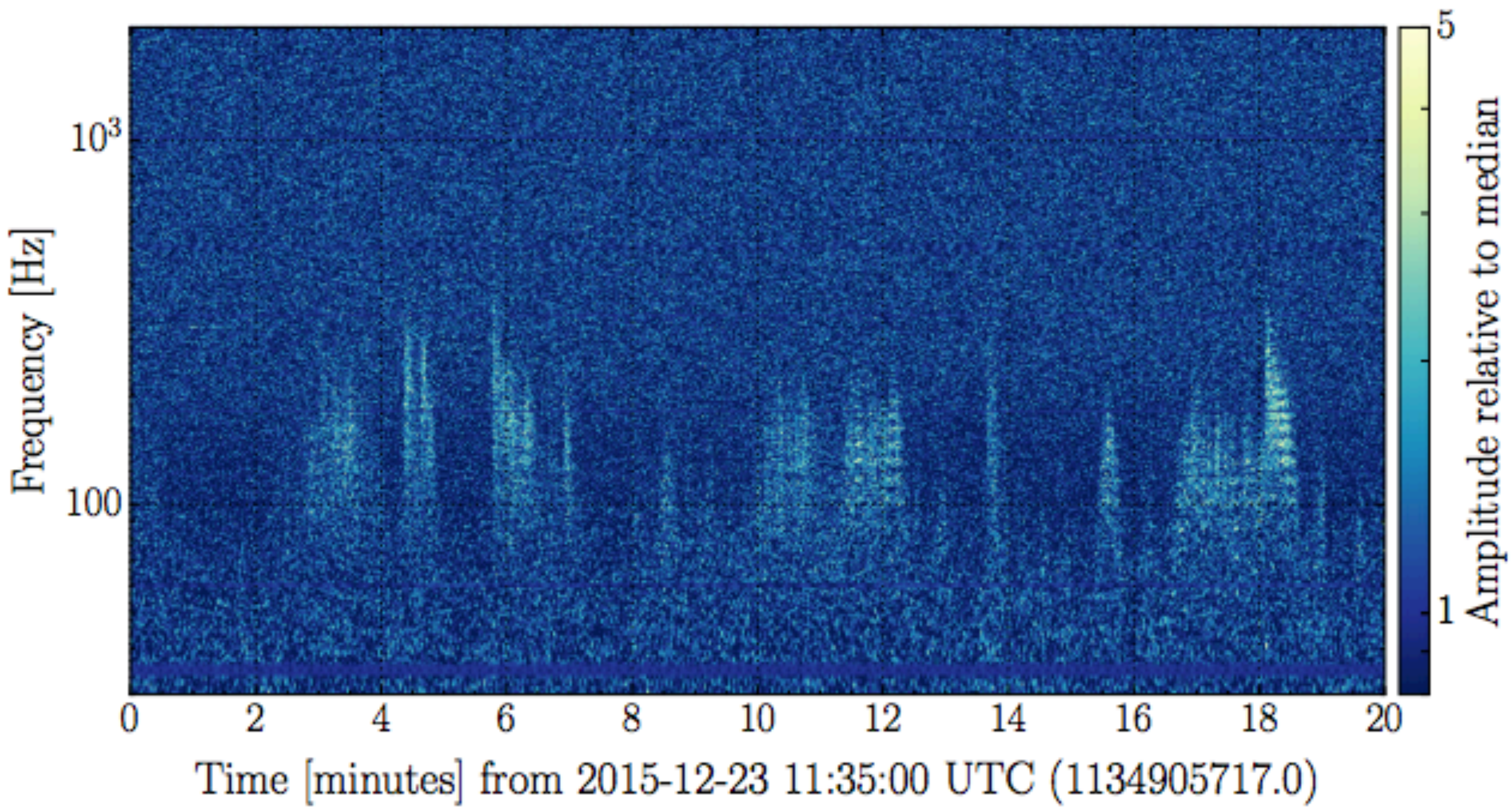
D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

https://www.zooniverse.org/projects/zooniverse/gravity-spy



In reality…GW data also contains instrumental and environmental artefacts

The data can also be non-stationary - meaning that it varies with time.

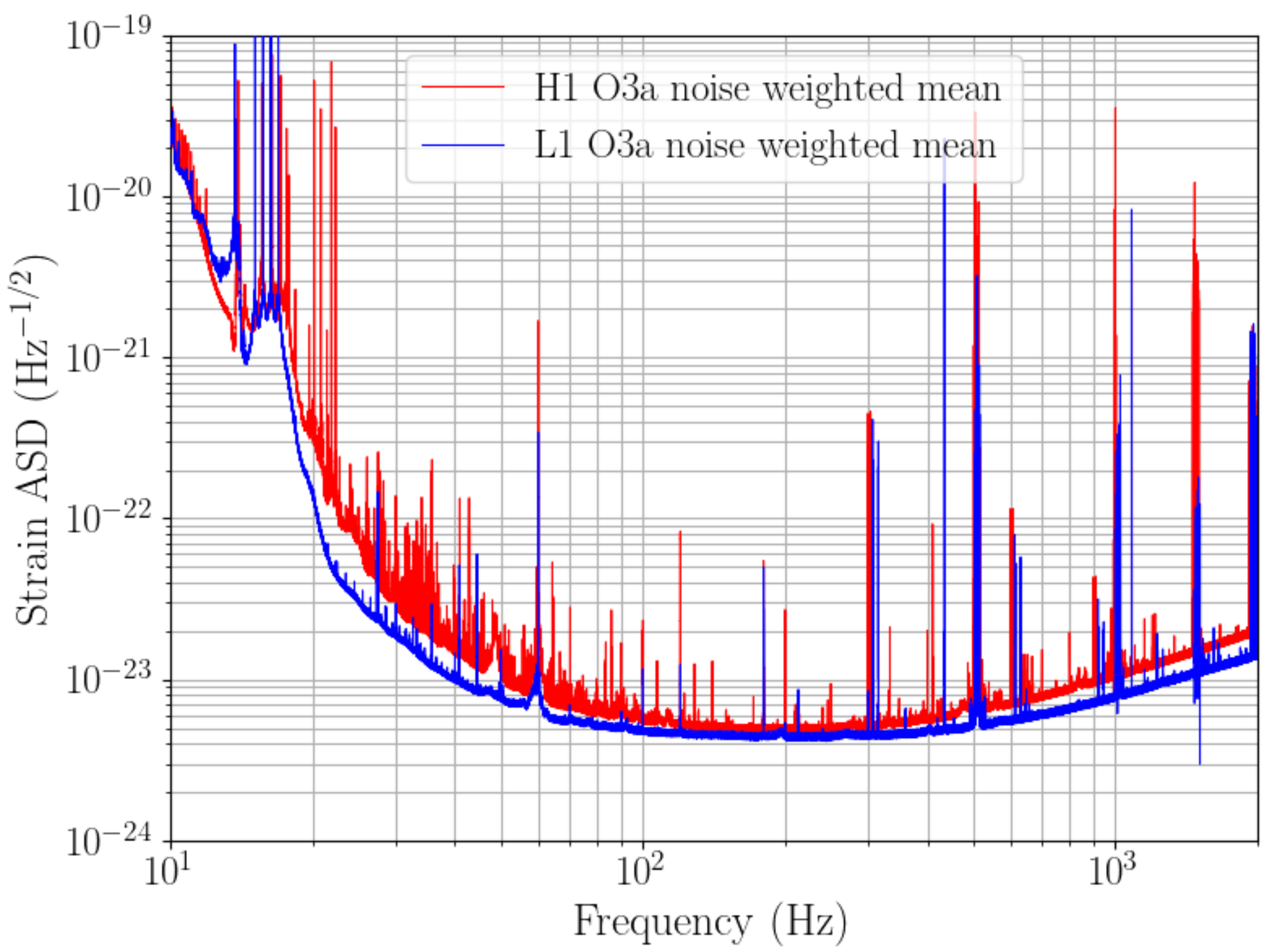


B.P. Abbott et al., *Effects of Data Quality Vetoes on a Search for Compact Binary Coalescences in Advanced LIGO's First Observing Run*, CQG 35, 065010 (2018)

In reality…GW data also contains instrumental and environmental artefacts

In the frequency domain is is clear to see many combs of lines in the data.

More information at: https://www.gw-openscience.org/O3/o3aspeclines/



Outline

- What is GWpy?

- What does GW data look like?
  - Time domain
  - Frequency domain
  - Time-frequency representation
  - Noise Subtraction

- Data Quality
  - Various forms of noise

- Identifying and Mitigating Sources of Noise
  - Data Quality Vetoes
  - Event Validation

- Data Quality Information
  - Detector Status Summary Pages

- Summary of resources and references

Data Quality Information Available in 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

Data Quality Information

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

Data Quality Information

DATA (Data Available): Failing this level indicates that LIGO/Virgo 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.*

Data Quality Information

DATA (Data Available): Failing this level indicates that LIGO/Virgo 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.*

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.

Data Quality Information

DATA (Data Available): Failing this level indicates that LIGO/Virgo 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.*

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):
 - Burst: 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.
 - CBC: Category not used

Data Quality Information

DATA (Data Available): Failing this level indicates that LIGO/Virgo 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.*

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.

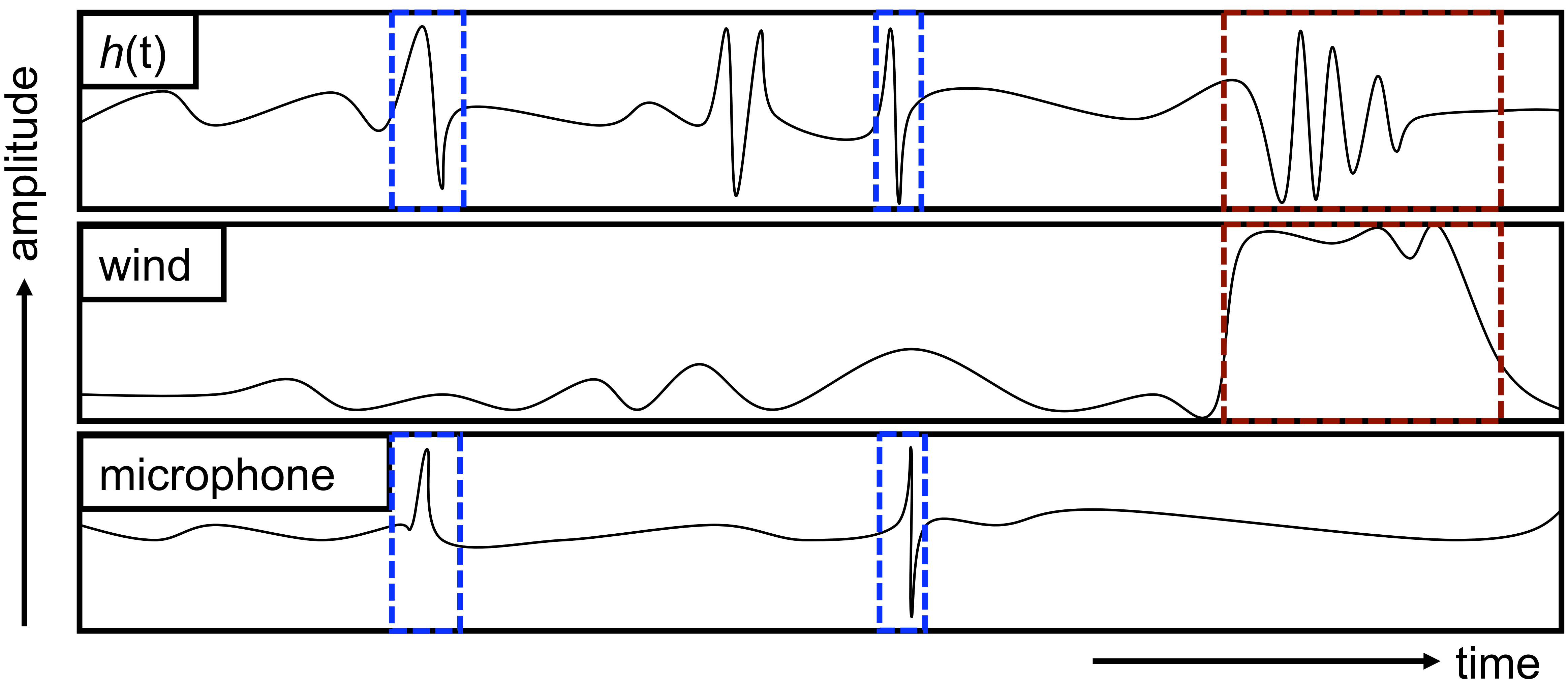
CAT3 (Category 3):
- Burst: Failing a data quality check at this category indicates times when there is **statistical coupling to the gravitational wave channel** which is not fully understood.
- CBC: Category not used

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.

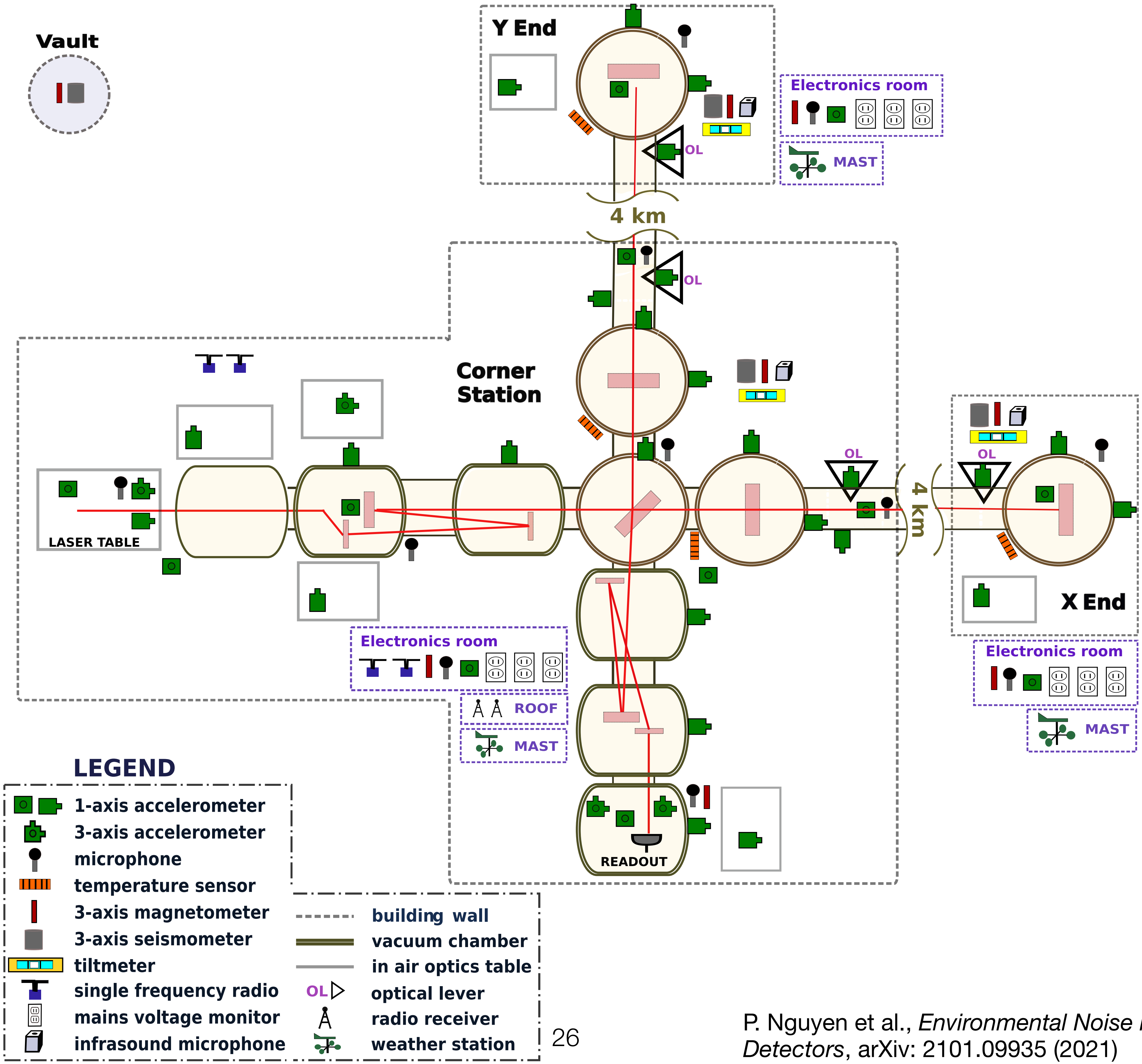
Auxiliary Channels

- We record over 200,000 channels per detector that monitor environment and detector behaviour
- We can use them to help track down and trace instrumental causes of glitches that pollute the searches.



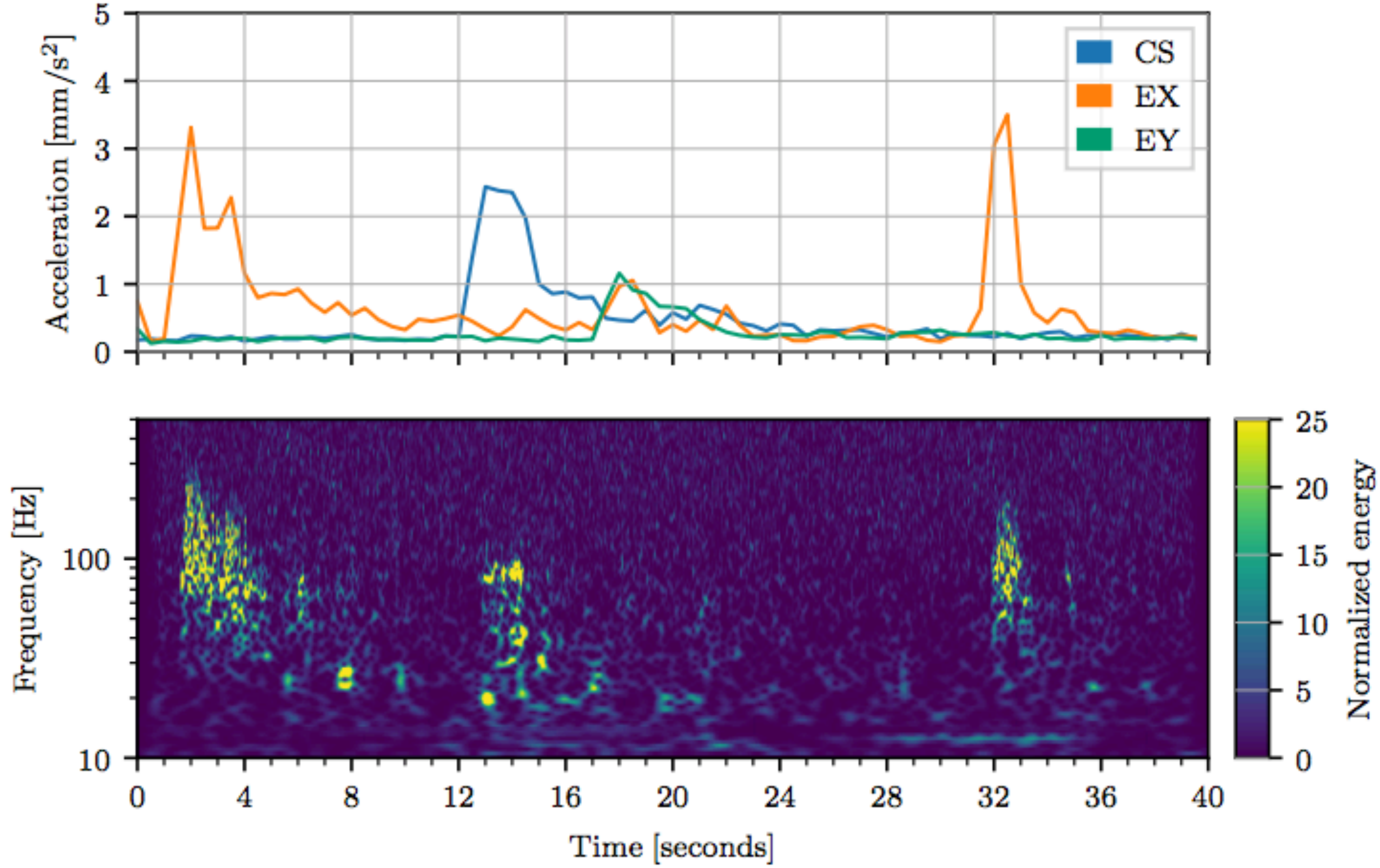
J.Smith et al., *A hierarchical method for vetoing noise transients in gravitational-wave detectors*, CQG **28** 235005 (2011)

Physical Environment Channels



P. Nguyen et al., *Environmental Noise in Advanced LIGO Detectors*, arXiv: 2101.09935 (2021)

Thunderstorms

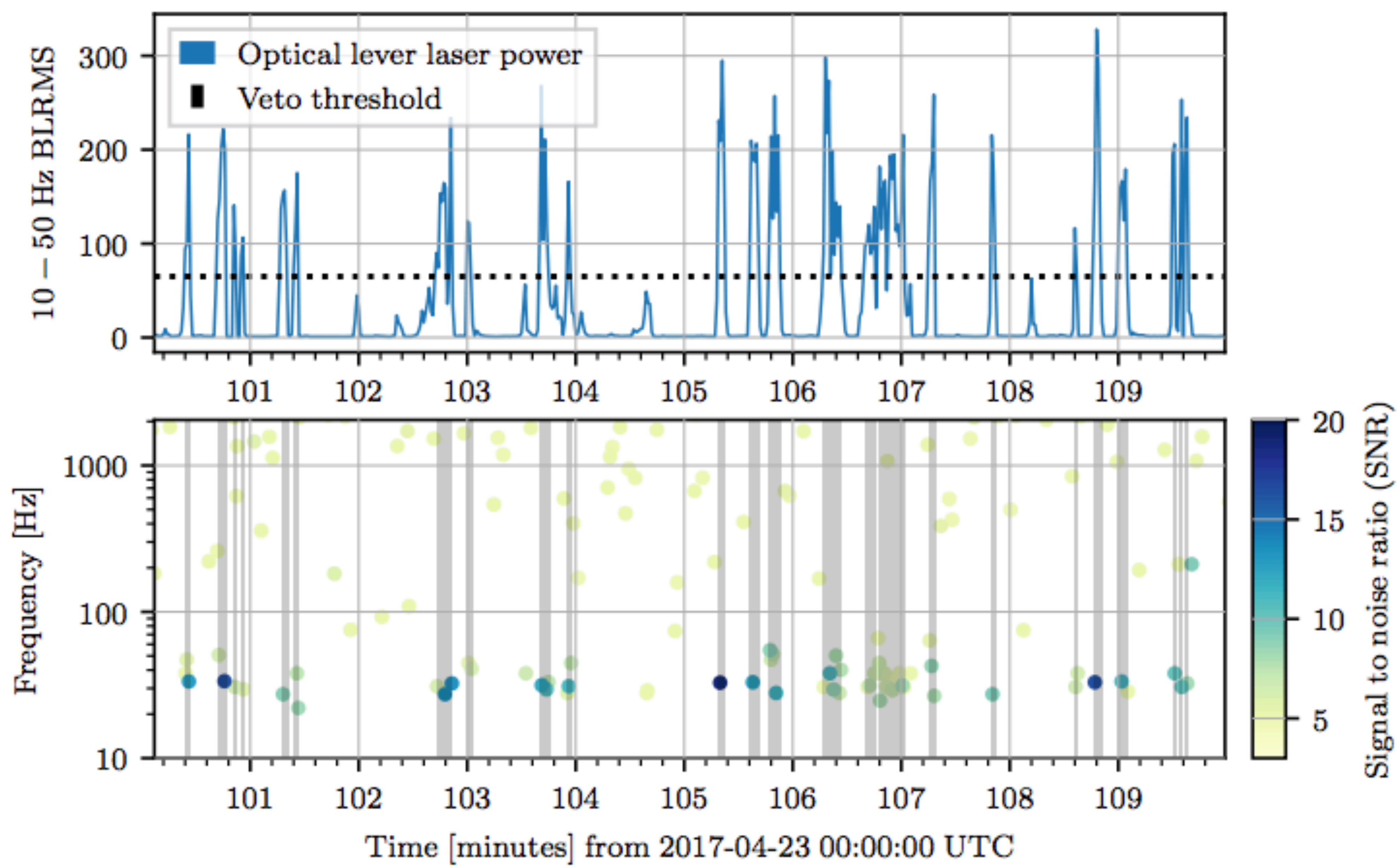


- Top: Data between 10-100 Hz from accelerometers located in the corner station (CS), End X station (EX) and End Y station (EY)

- Bottom: Spectrogram of the GW strain channel at the same time. Excess noise in the frequency range of 20 Hz to 200 Hz coincides with the thunderclaps, with intensity depending on the thunder's location.

D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

Example of a data quality veto in O2

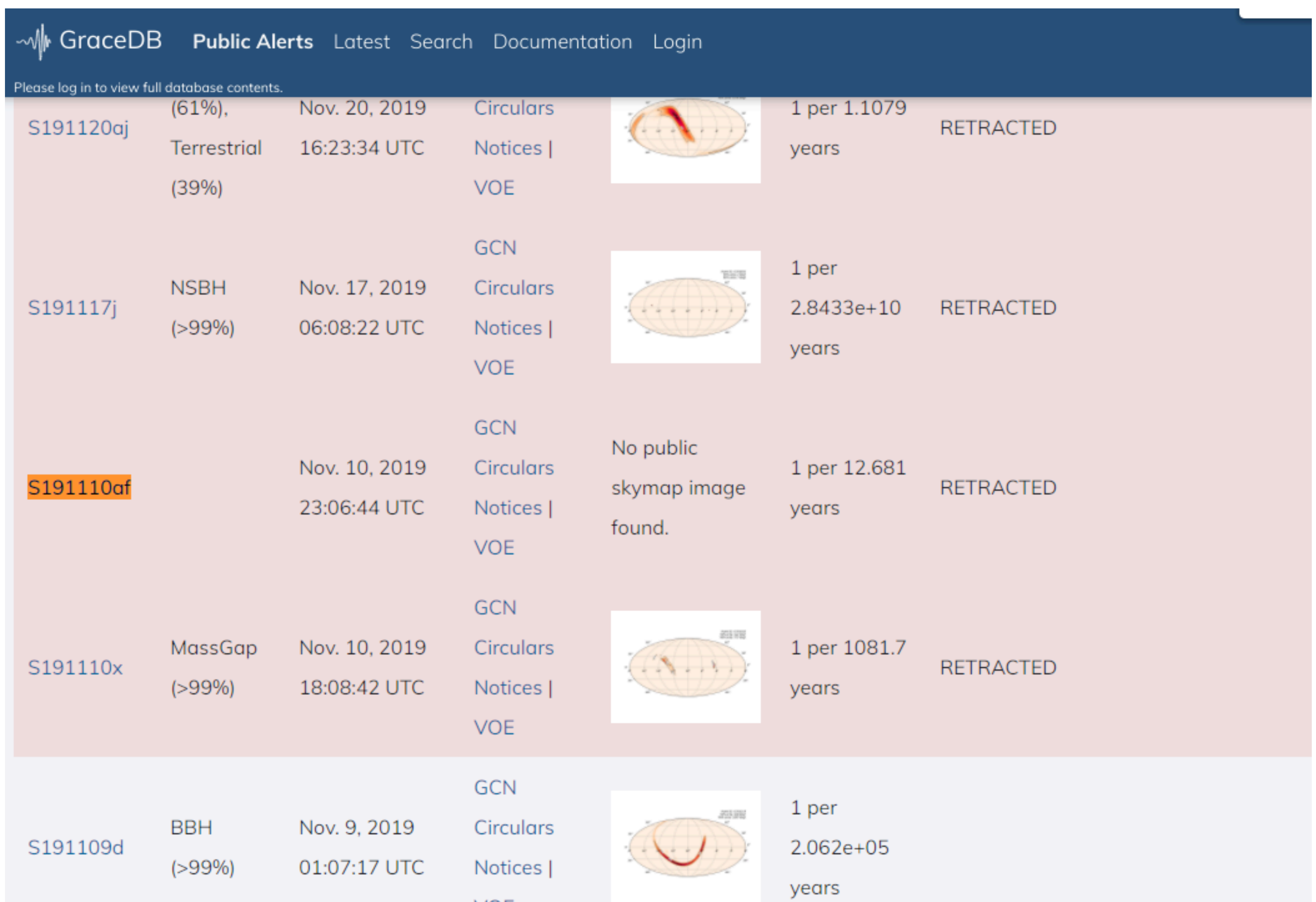


D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

Data quality of individual events

Evaluation of the data quality around an event is important to:

- identify a clear instrumental origin and issue a retraction


GraceDB Public Alerts Latest Search Documentation Login

Please log in to view full database contents.

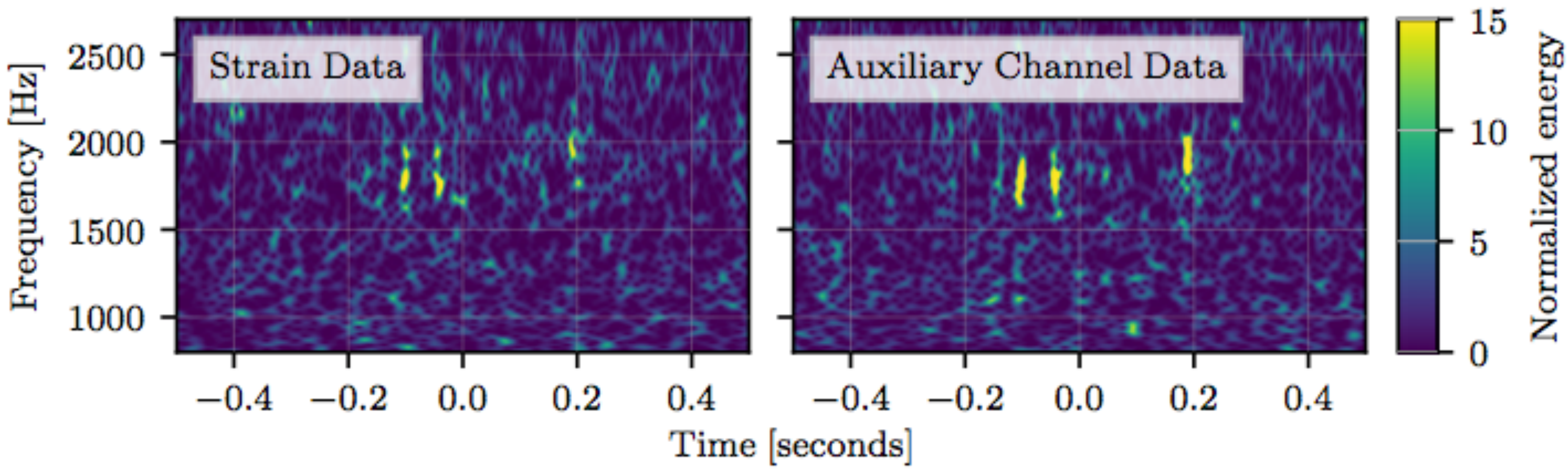
S191120aj	(61%), Terrestrial (39%)	Nov. 20, 2019 16:23:34 UTC	Circulars Notices | VOE		1 per 1.1079 years	RETRACTED
S191117j	NSBH (>99%)	Nov. 17, 2019 06:08:22 UTC	GCN Circulars Notices | VOE		1 per 2.8433e+10 years	RETRACTED
S191110af		Nov. 10, 2019 23:06:44 UTC	GCN Circulars Notices | VOE	No public skymap image found.	1 per 12.681 years	RETRACTED
S191110x	MassGap (>99%)	Nov. 10, 2019 18:08:42 UTC	GCN Circulars Notices | VOE		1 per 1081.7 years	RETRACTED
S191109d	BBH (>99%)	Nov. 9, 2019 01:07:17 UTC	GCN Circulars Notices | VOE		1 per 2.062e+05 years	

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

S191110af

- Potential Burst Source in LIGO and Virgo
- Looking at the LIGO-Hanford data, there was a clear correlation between an auxiliary channel (i.e. not sensitive to GWs) and the gravitational-wave strain channel

  - Similar morphology between the two channels
  - Origin of this event is instrumental (from the output mode cleaner) rather than astrophysical



D. Davis et al., *LIGO Detector Characterization in the Second and Third Observing Runs*, arXiv: 2101.11673 (2021)

Data quality of individual events

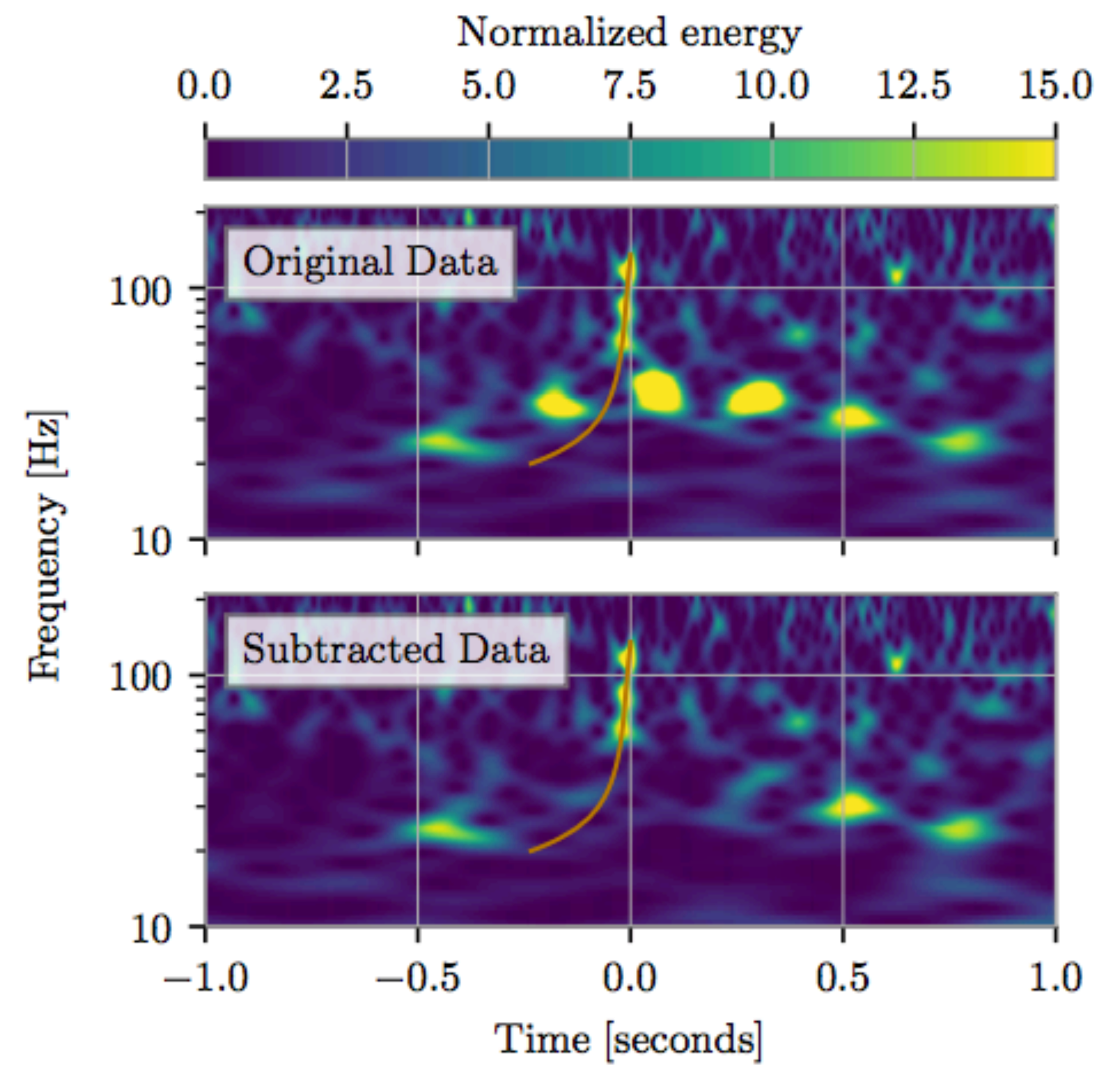
Evaluation of the data quality around an event is important to:

- identify a clear instrumental origin and issue a retraction
- rule out an instrumental origin
  (i.e. all GW events that have
  been published)

Data quality of individual events

Evaluation of the data quality around an event is important to:

- identify a clear instrumental origin and issue a retraction
- rule out an instrumental origin (i.e. all GW events that have been published)
- identify if any instrumental noise needs to be mitigated before an analysis to determine the GW parameters is completed
  - i.e. glitch subtraction around candidate events.



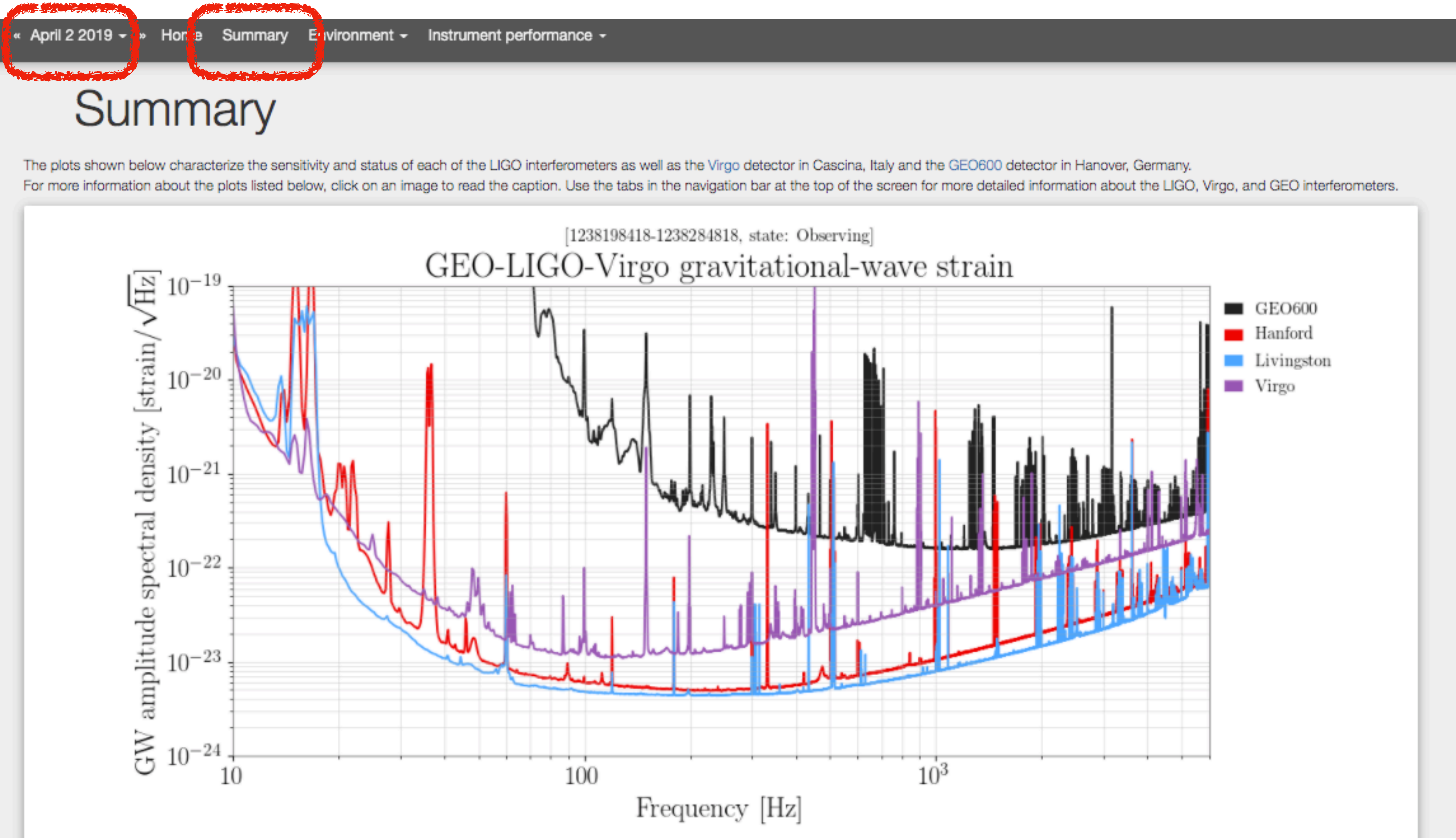
Scattered light was present around the event GW190701_203306. Despite the overlap, the excess power from the glitch is successfully modelled and subtracted

R. Abbott et al., *GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*, arXiv: 2010.14527 (2020)

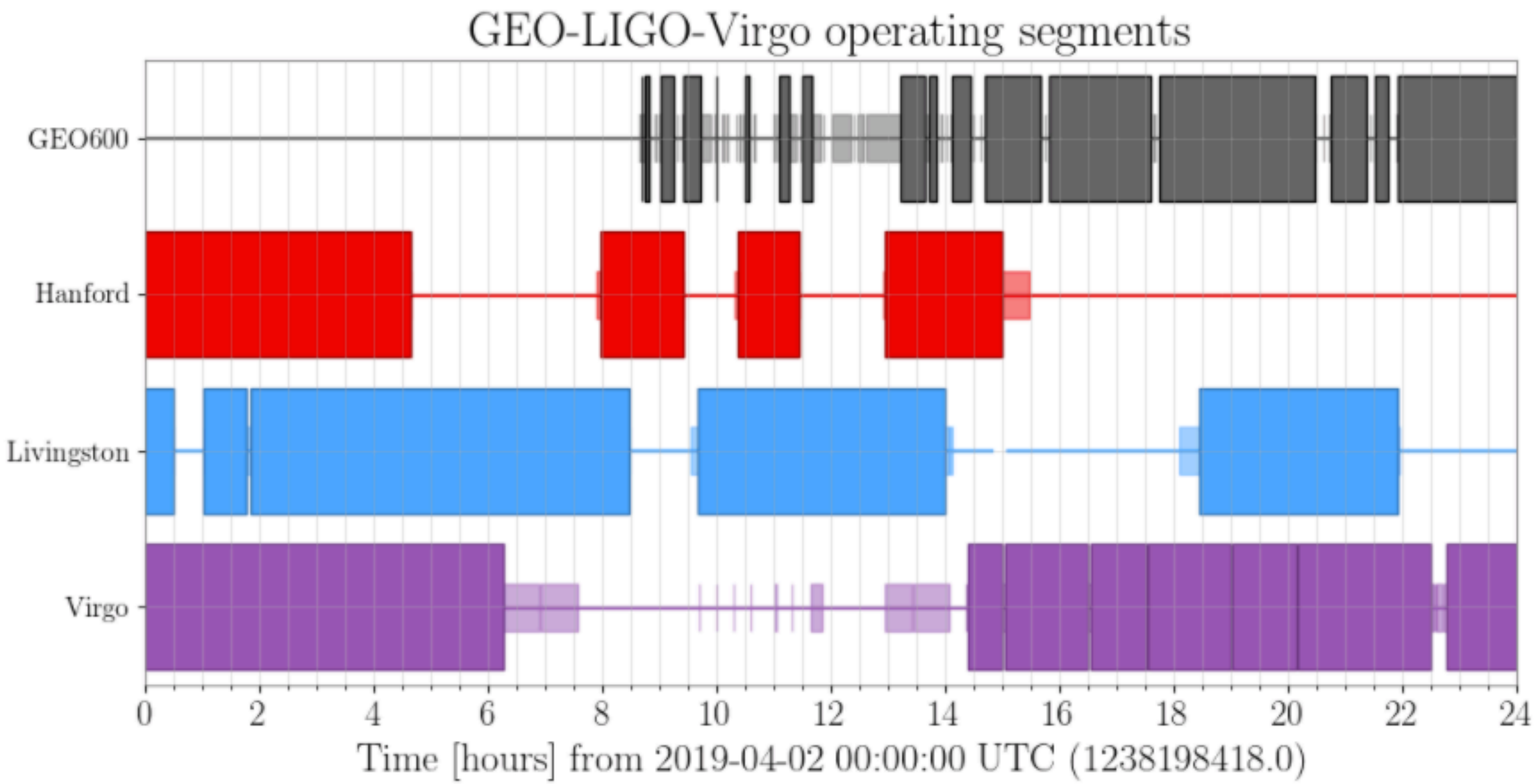
Outline

- What is GWpy?
- What does GW data look like?
  - Time domain
  - Frequency domain
  - Time-frequency representation
  - Noise Subtraction
- Data Quality
  - Various forms of noise
- Identifying and Mitigating Sources of Noise
  - Data Quality Vetoes
  - Event Validation
- Data Quality Information
  - Detector Status Summary Pages
- Summary of resources and references

Detector Status:
https://www.gw-openscience.org/detector_status/

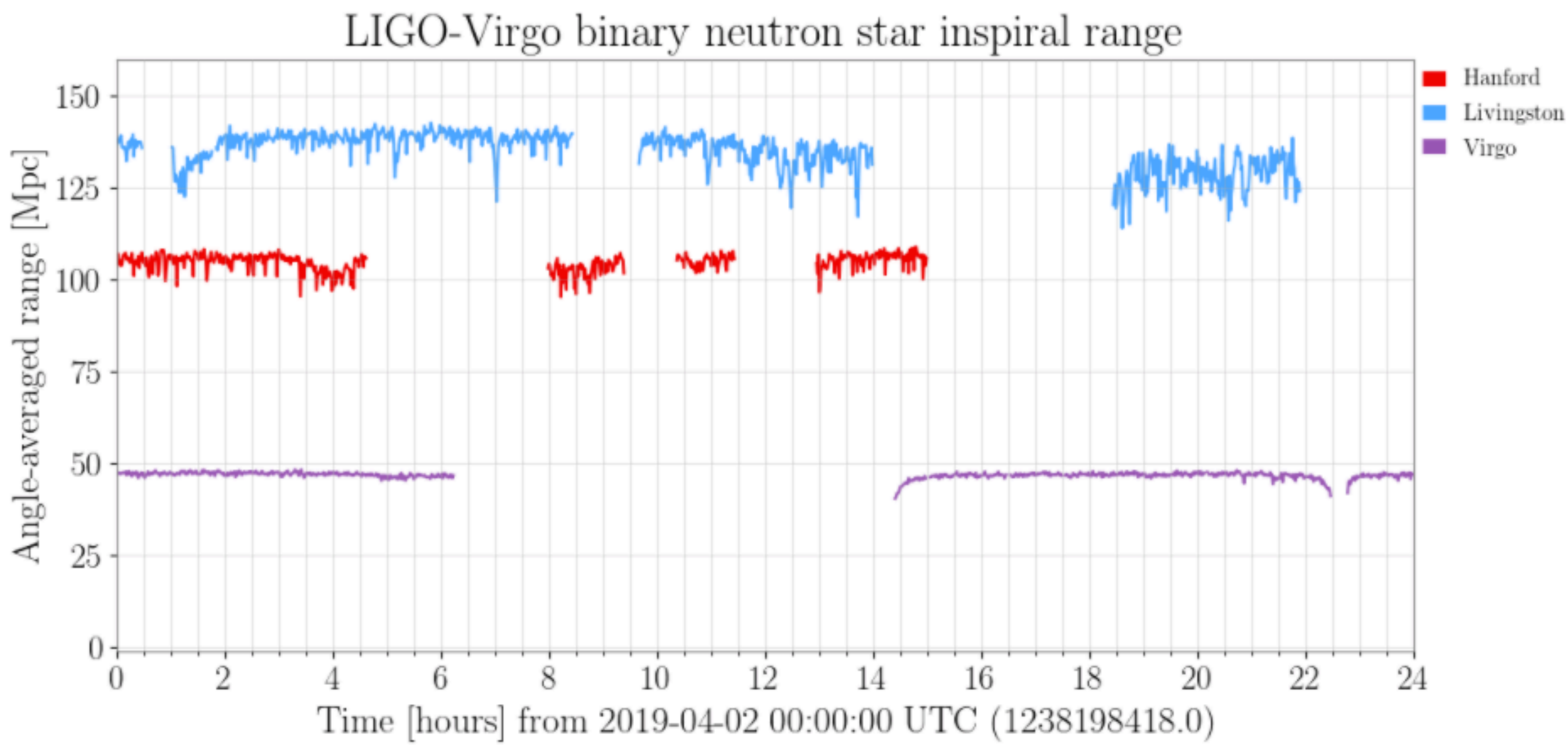


https://www.gw-openscience.org/detector_status/day/20190402/



This plot shows the instrument status for each detector, represented by horizontal bars, over the course of the day. The tall opaque blocks indicate times when the interferometer is intended to be in observing mode and no tests are currently being performed. The narrow, transparent blocks indicate times when the interferometer was in an operational state, but not in observing mode. This state often means that testing or commissioning is in progress.

https://www.gw-openscience.org/detector_status/day/20190402/



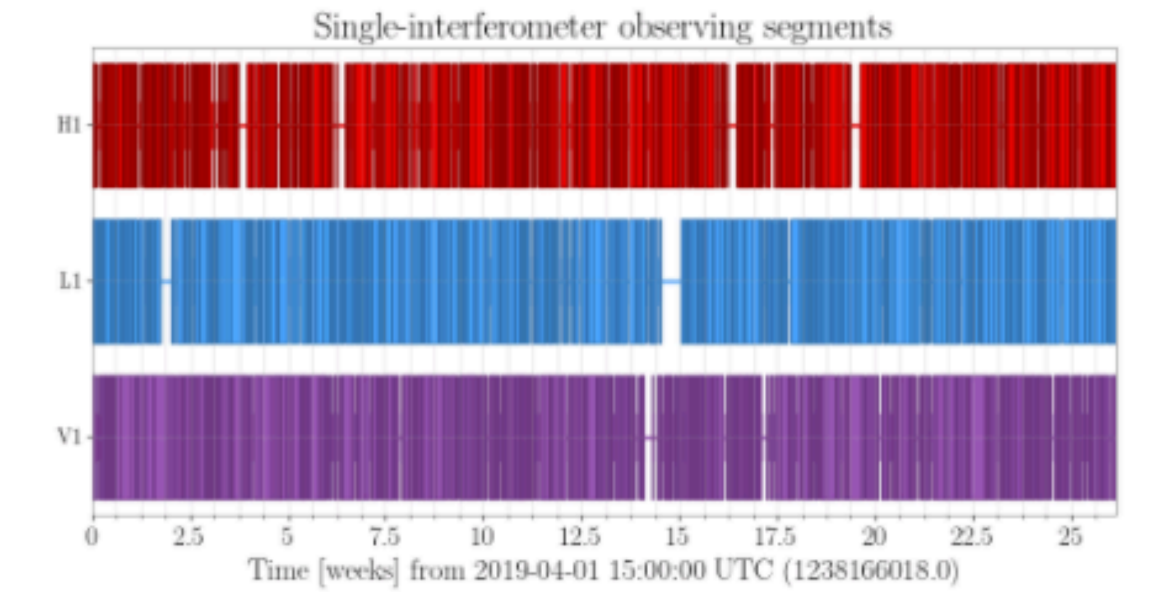
This plot shows average sensitivity of the Hanford, Livingston, and Virgo detectors to gravitational wave signals from binary neutron star mergers, as measured in megaparsecs (where 1 megaparsec = 3,262,000 light-years), averaged over all sky positions and source orientations. The value reported in this plot is determined by comparing the noise curve, as seen averaged over the entire day in the first plot, to a predicted gravitational wave signal from a binary neutron star merger. The sharp drops in this curve are typically due to transient noise in the interferometer limiting sensitivity to gravitational waves. The sharp upward peaks in the Virgo curve are artifacts caused by data dropouts.

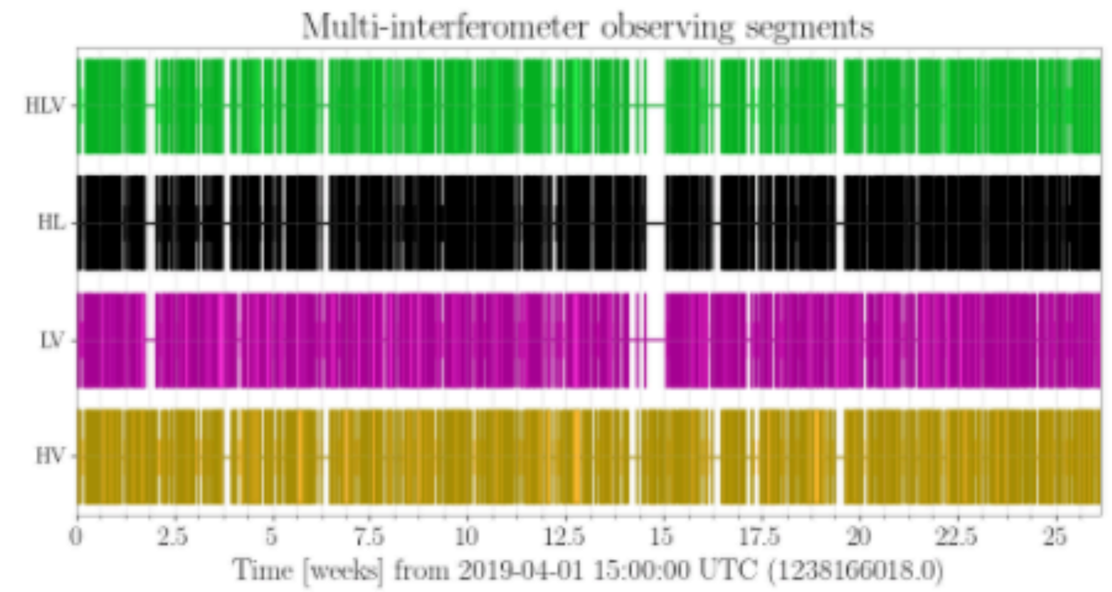
https://www.gw-openscience.org/detector_status/O3a/

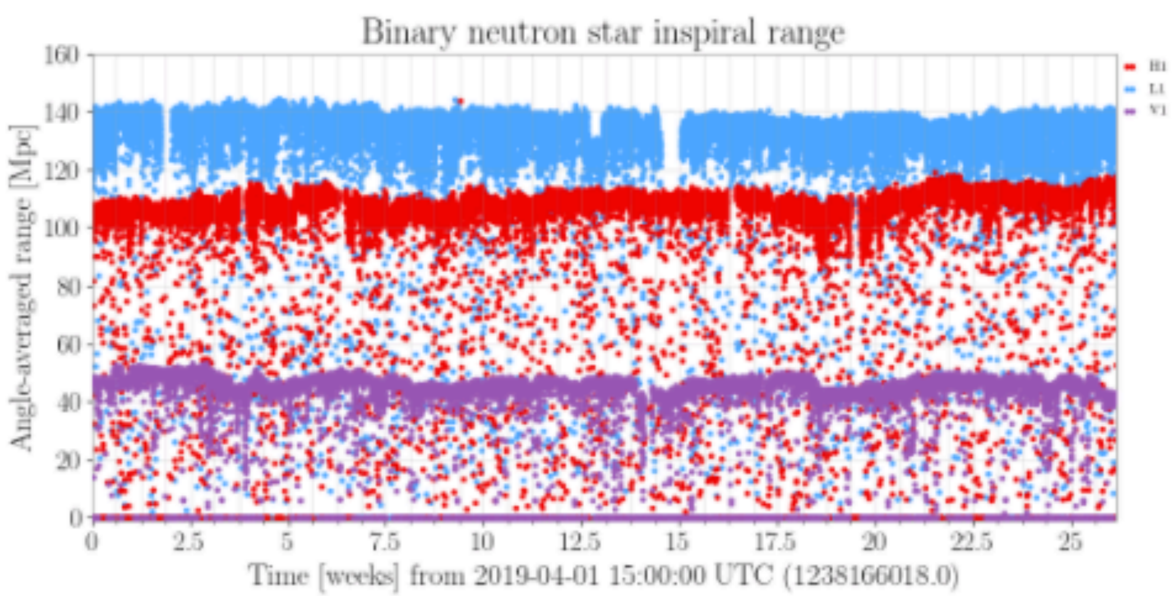
1238166018-1253977218 Home Summary

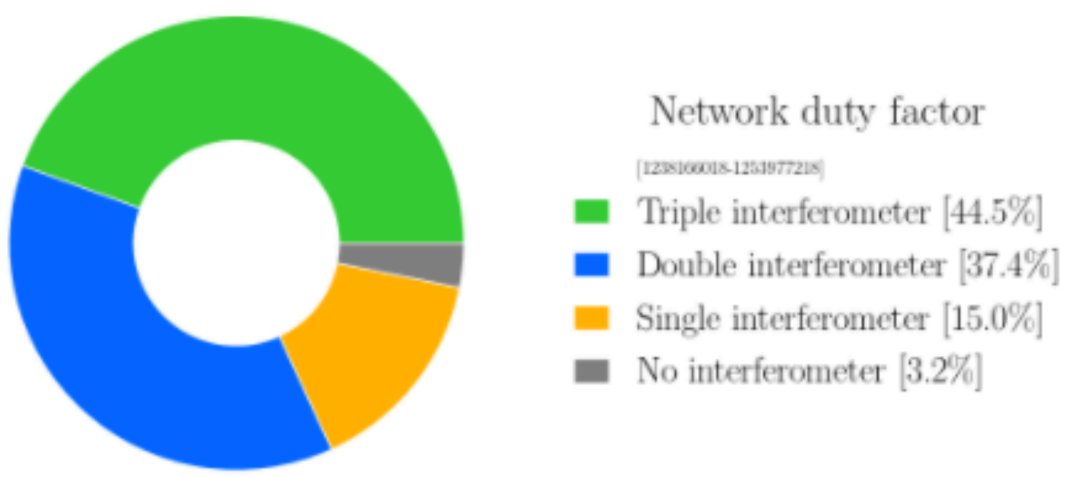
O3a Summary

The plots shown below characterize the sensitivity and status of each of the LIGO and Virgo interferometers during the first half of the third observing run, known as O3a, which began on April 1, 2019 and ended on October 1, 2019. For more information about the plots listed below, click on an image to read the caption.









Resources

For glitches:
GW150914 Detector Characterization paper: arXiv 1602.03844
O2/O3 LIGO Detector Chracterization paper: arXiv: 2101.11673
O3 Virgo paper: coming soon!
Gravity Spy: gravityspy.org

For lines:
O1/O2 lines paper: arXiv 1801.07204
O3a lines calico on GWOSC: https://www.gw-openscience.org/O3/o3aspeclines/
O2 lines catalog on the GWOSC: https://www.gw-openscience.org/o2speclines/

Data Quality around events: GWTC-2 paper: arXiv: 2010.14527

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

O3a Data Set technical Details: https://www.gw-openscience.org/O3/o3a_details/

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

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

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