Data quality in gravitational-wave detectors

Hirotaka Yuzurihara

Gravitational Wave Open Data Workshop #7 (2024)

Acknowledgements

- Many slides are adapted from data quality presentations made by Ronaldas Macas, Laura Nuttall, Marissa Walker, and Jess McIver.

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

Outline

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

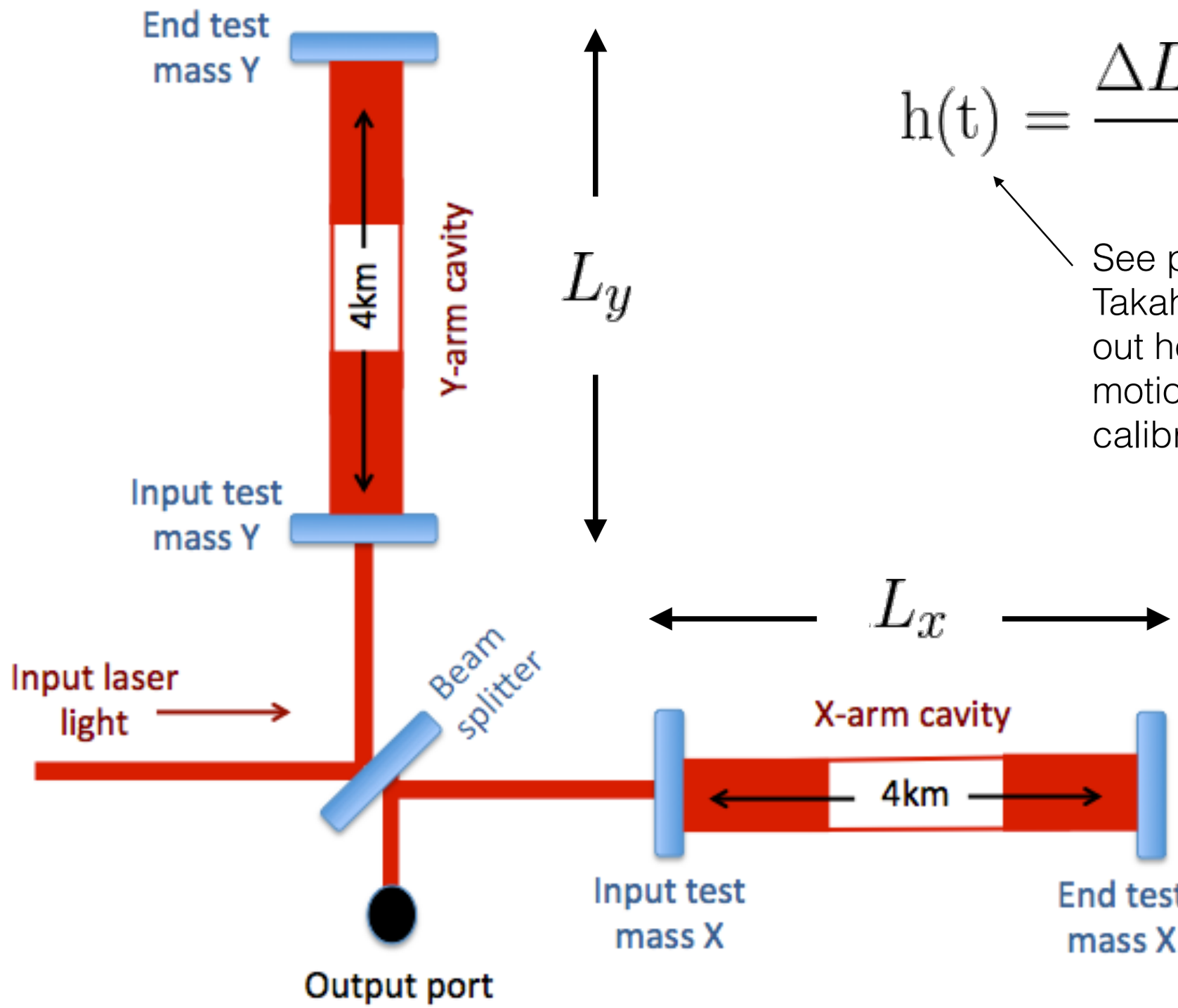
- Data quality: noise artifacts in strain data
  - Glitch
  - Lines

- Mitigating noise artifacts
  - Data quality vetoes
  - Analysis dependent mitigation
  - Event validation

- Data Quality Information
  - Summary page

- Reference

What is strain data, *h(t)* ?

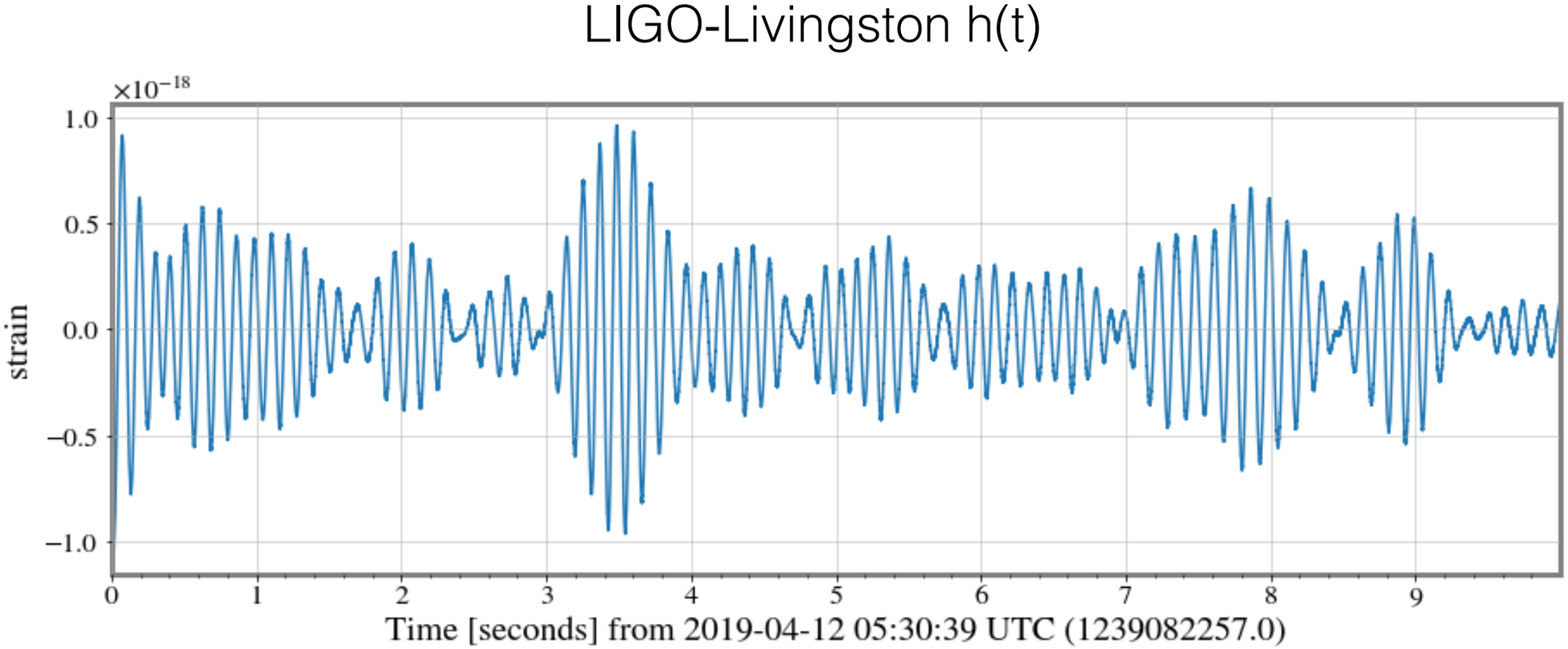


$$\mathrm{h(t)} = \frac{\Delta L_y - \Delta L_x}{L}$$

See previous talk by Takahiro Sawada to find out how the differential arm motion control signal is calibrated into h(t)

Slide by J McIver

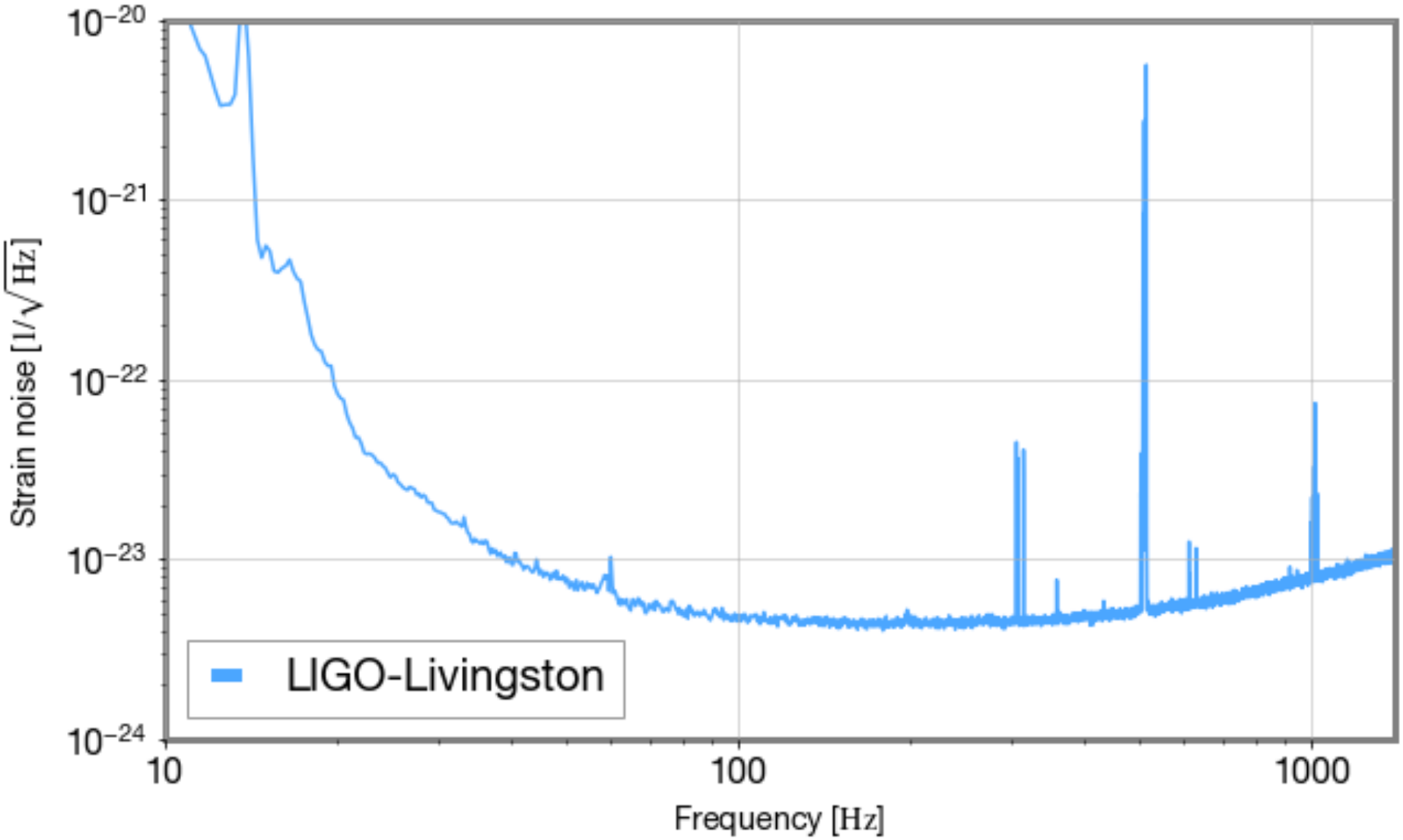
What does strain data look like?



- h(t) sampling rate for open data: 16384 or 4096 Hz

https://github.com/gw-odw/odw-2024/blob/main/Tutorials/Day_1/Tuto_1.2_Open_Data_access_with_GWpy.ipynb

Strain data in the frequency domain



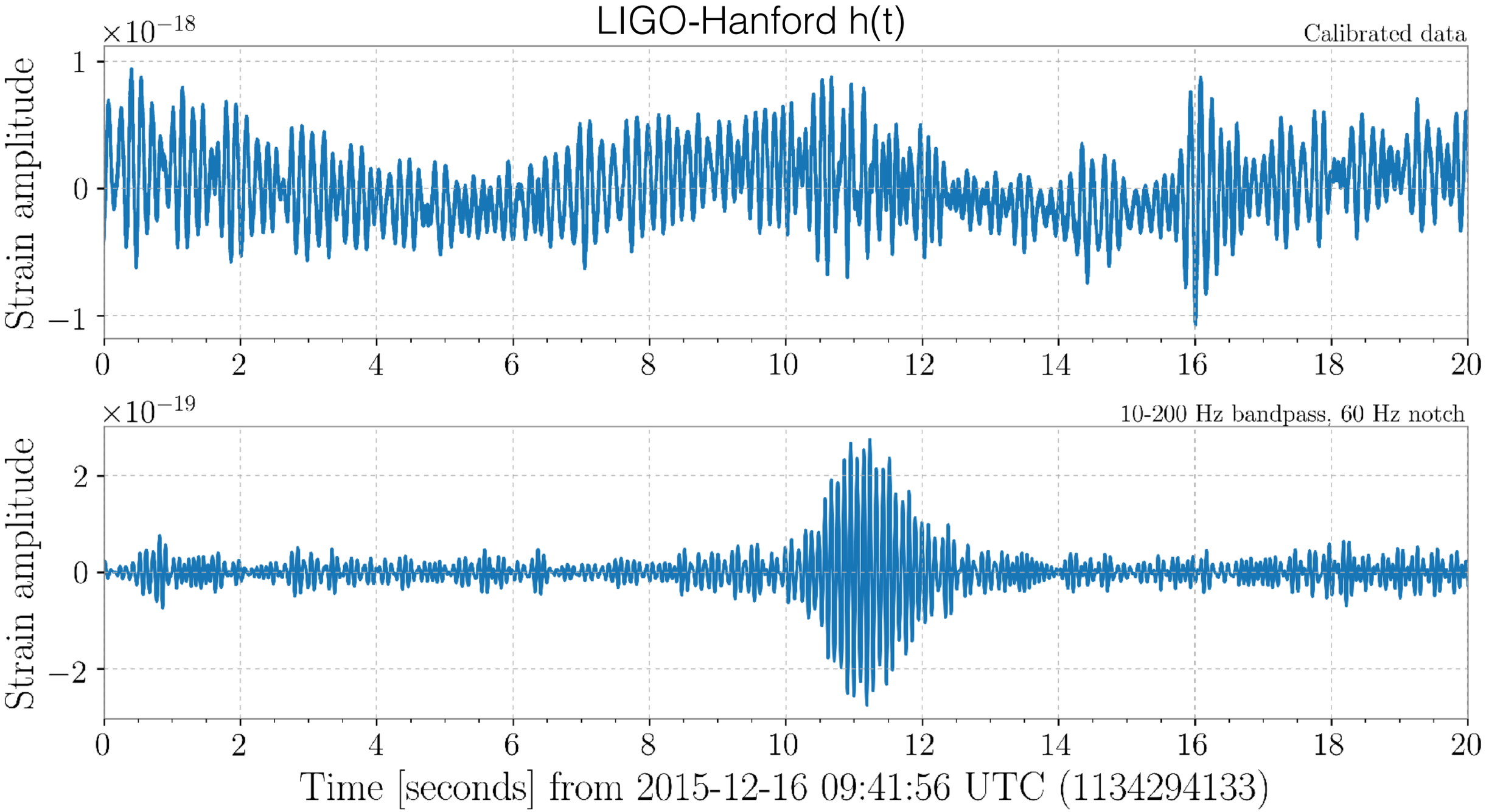
- Amplitude spectrum density (ASD) of detector sensitivity
  - Median
  - Mean

$$\mathrm{PSD}(f) = \lim_{T\to\infty} \frac{2|X(f)|^2}{T}$$

$$\mathrm{ASD}(f) = \sqrt{\mathrm{PSD(f)}}$$

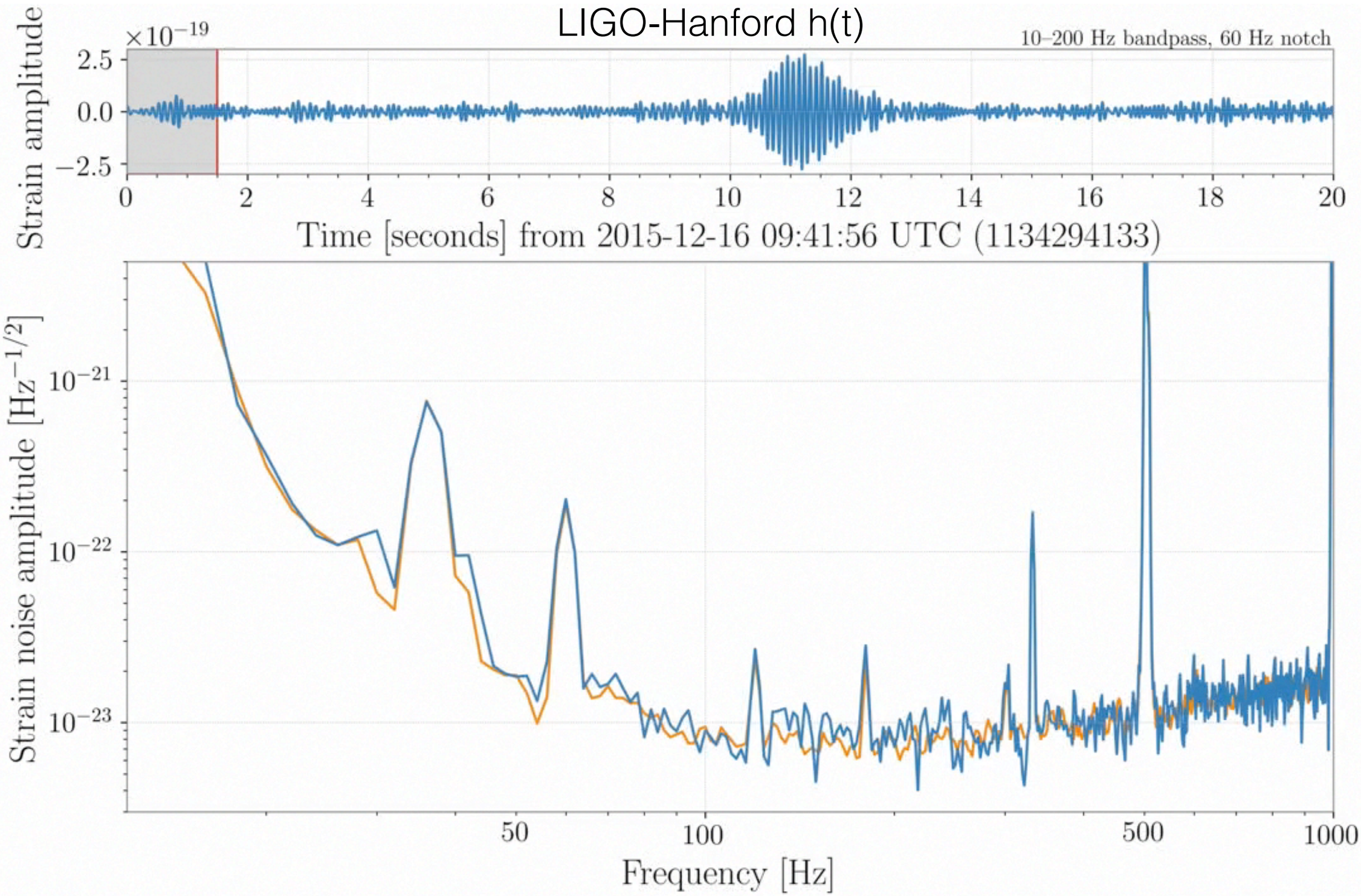
https://github.com/gw-odw/odw-2024/blob/main/Tutorials/Day_1/Tuto_1.2_Open_Data_access_with_GWpy.ipynb

What does strain data look like?



Made with GWpy by Duncan Macleod

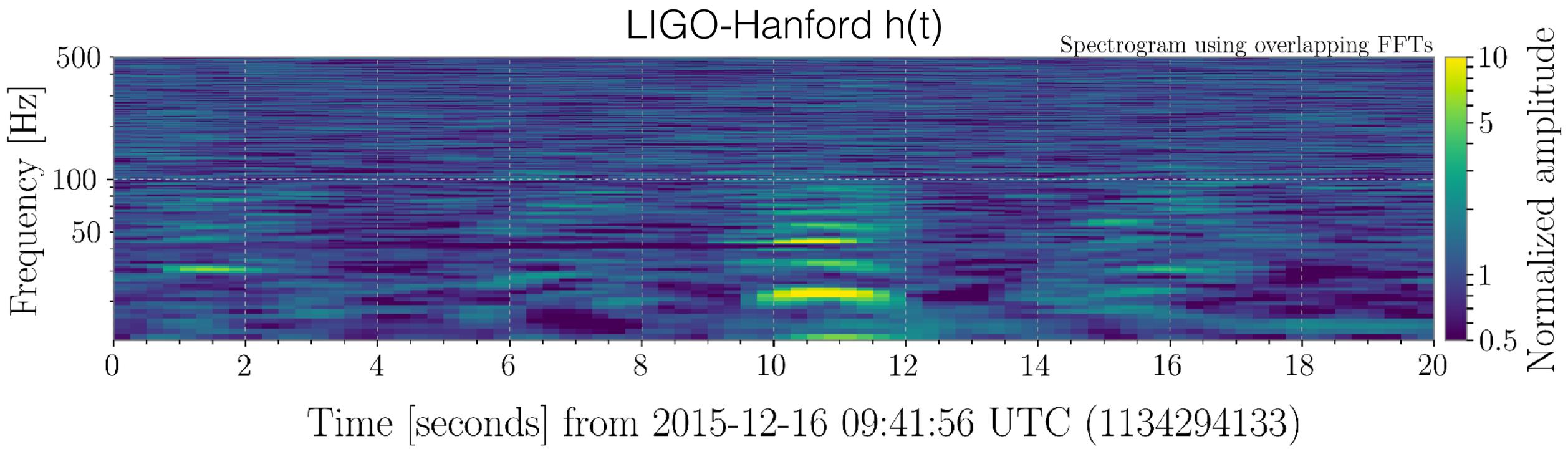
Strain data in the frequency domain



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

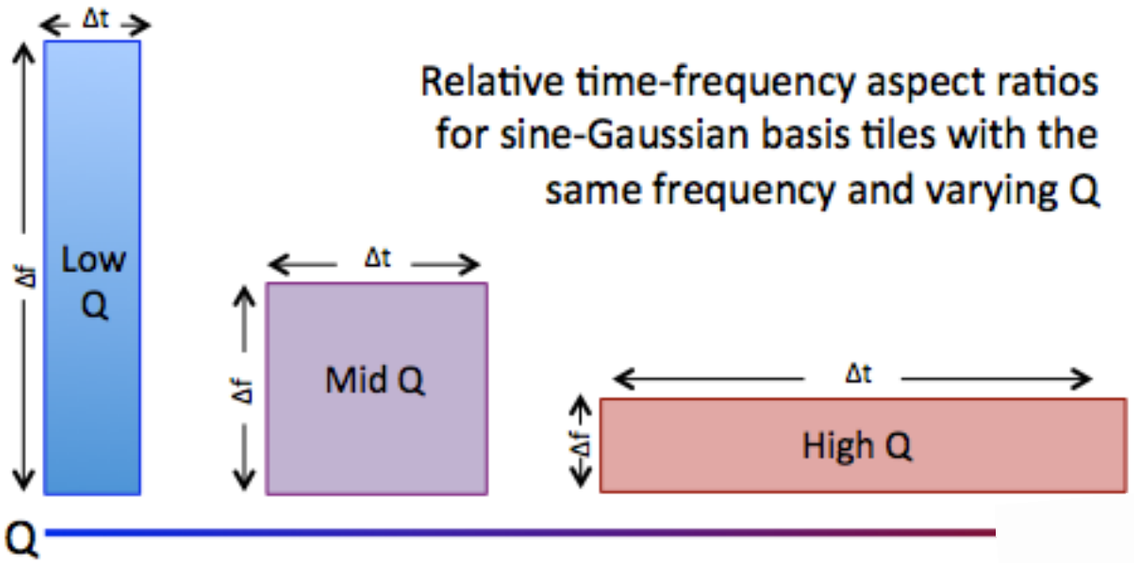
0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

Time-frequency spectrogram



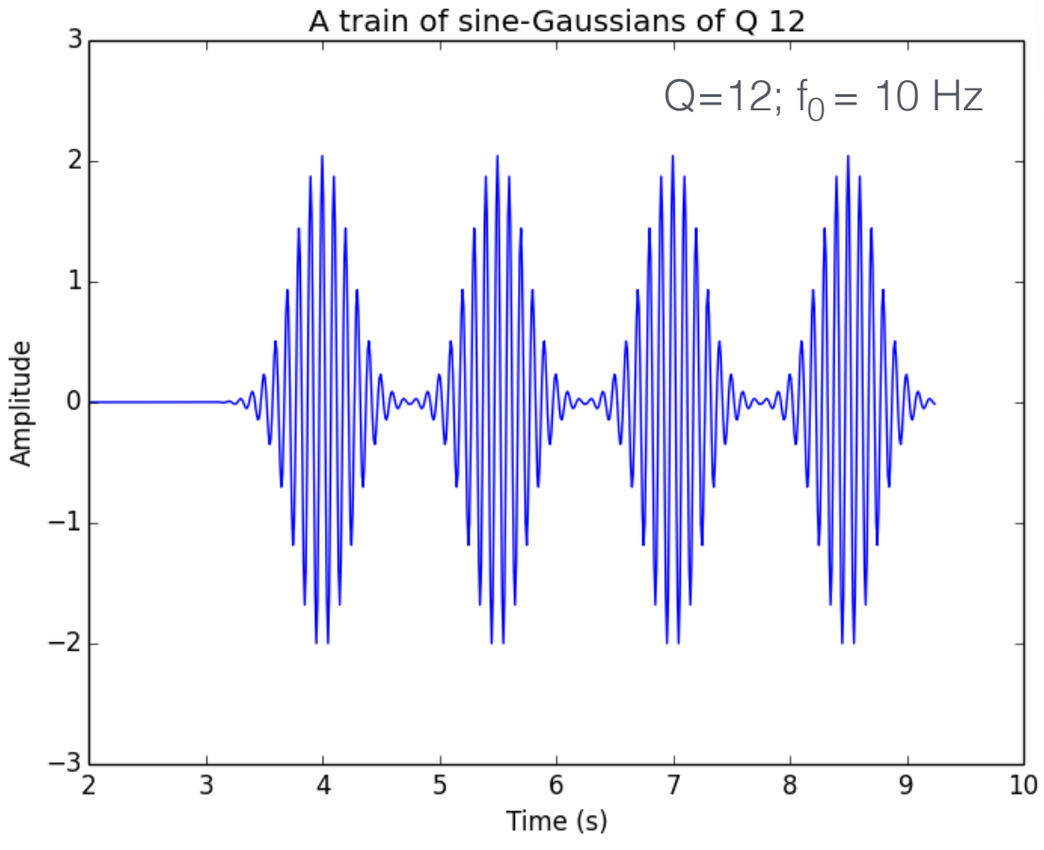
Made with GWpy by Duncan Macleod

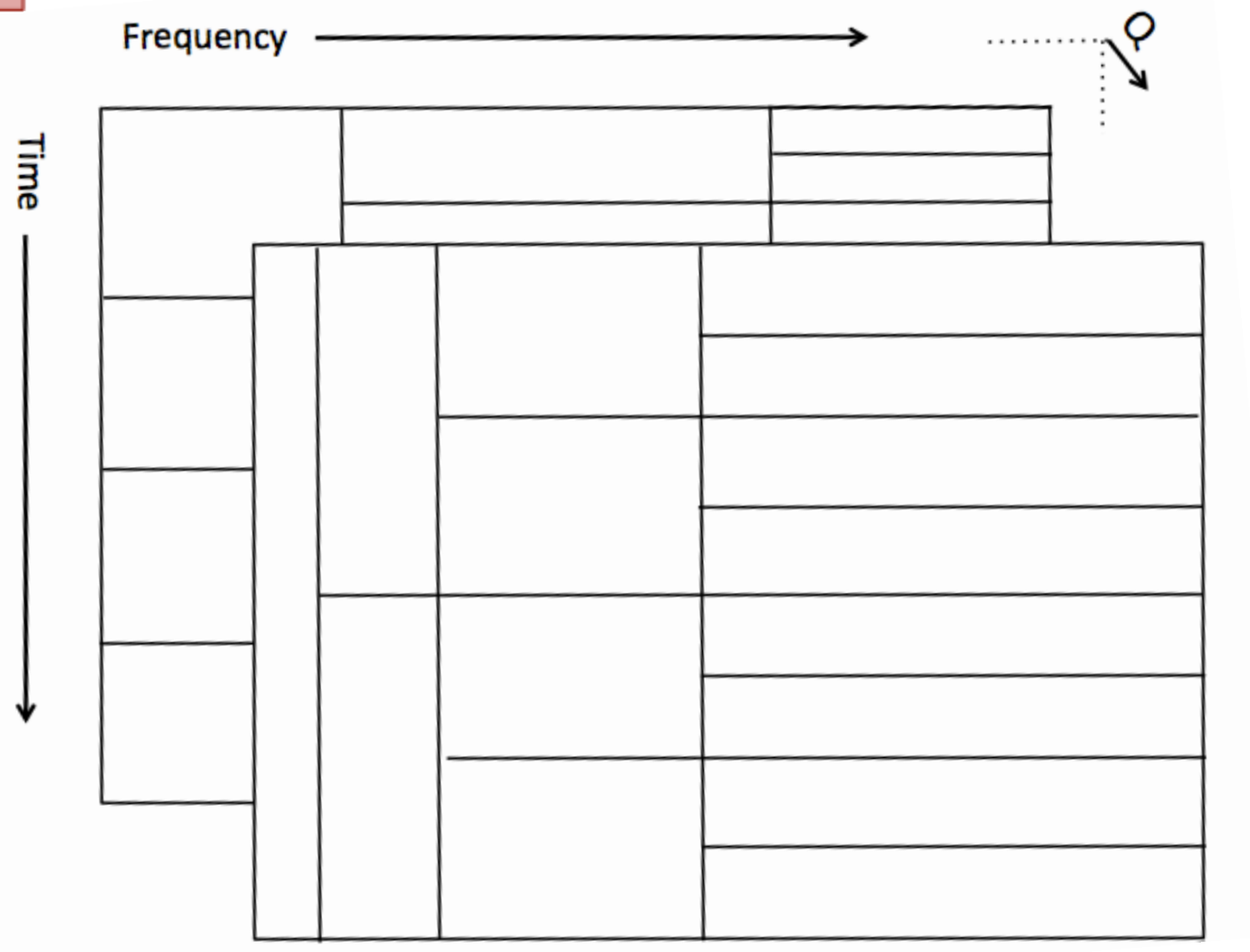
Q transform



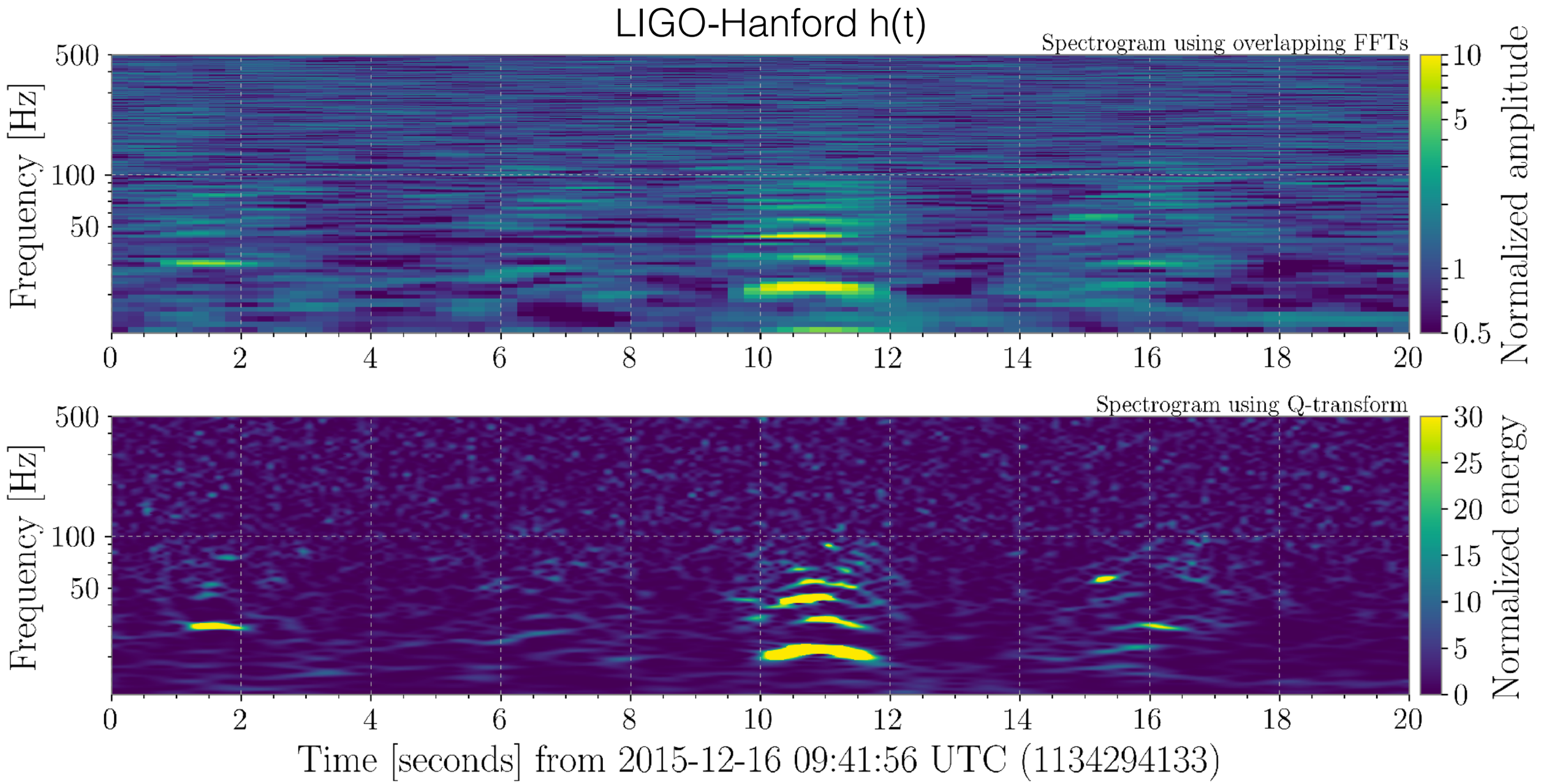
S. Chatterji et al. CQG (2010)
Images: McIver

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





Time-frequency spectrograms



Made with GWpy by Duncan Macleod

Noise Subtraction for O3 dataset

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.

For details, see https://gwosc.org/O3/o3_details/

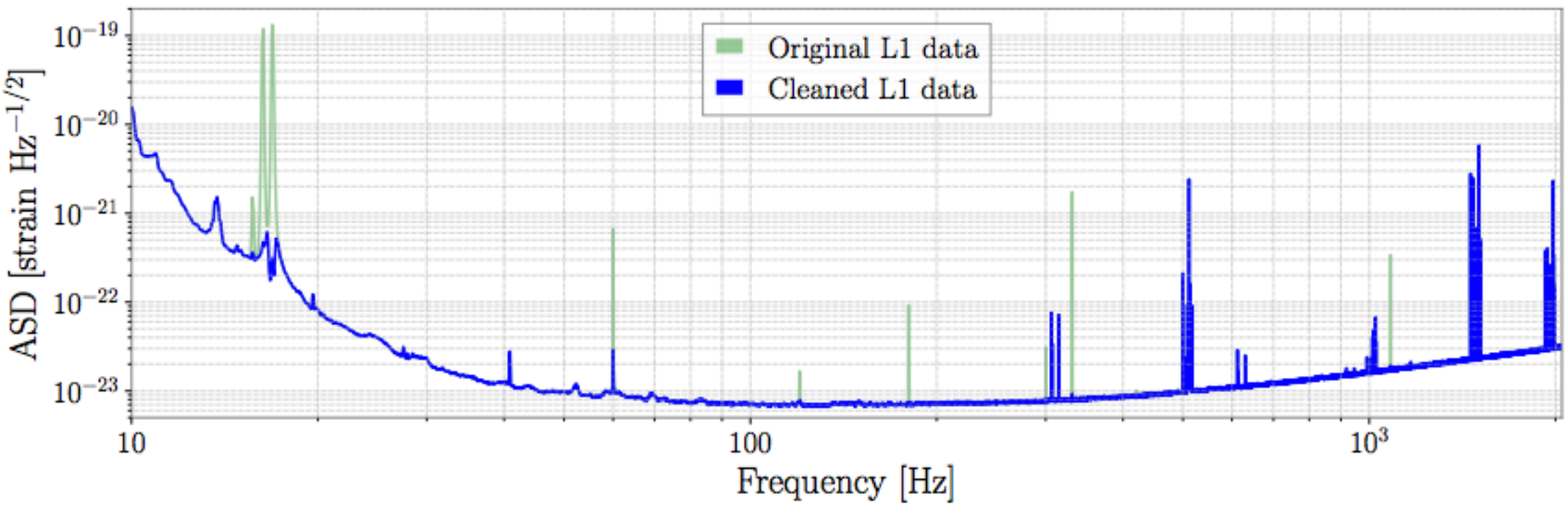


Image: D. Davis et al., *Improving the sensitivity of Advanced LIGO using noise subtraction*, CQG **36** 055011 (2019)

Outline

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

- Data quality: noise artifacts in strain data
  - Glitch
  - Lines

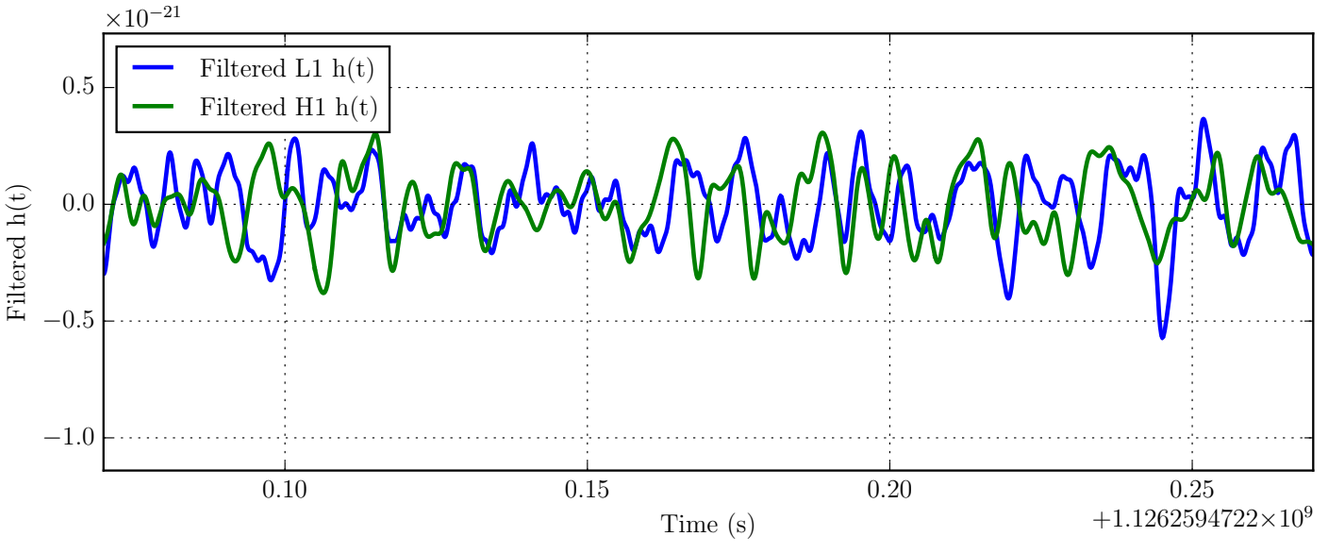
- Mitigating noise artifacts
  - Data quality vetoes
  - Analysis dependent mitigation
  - Event validation

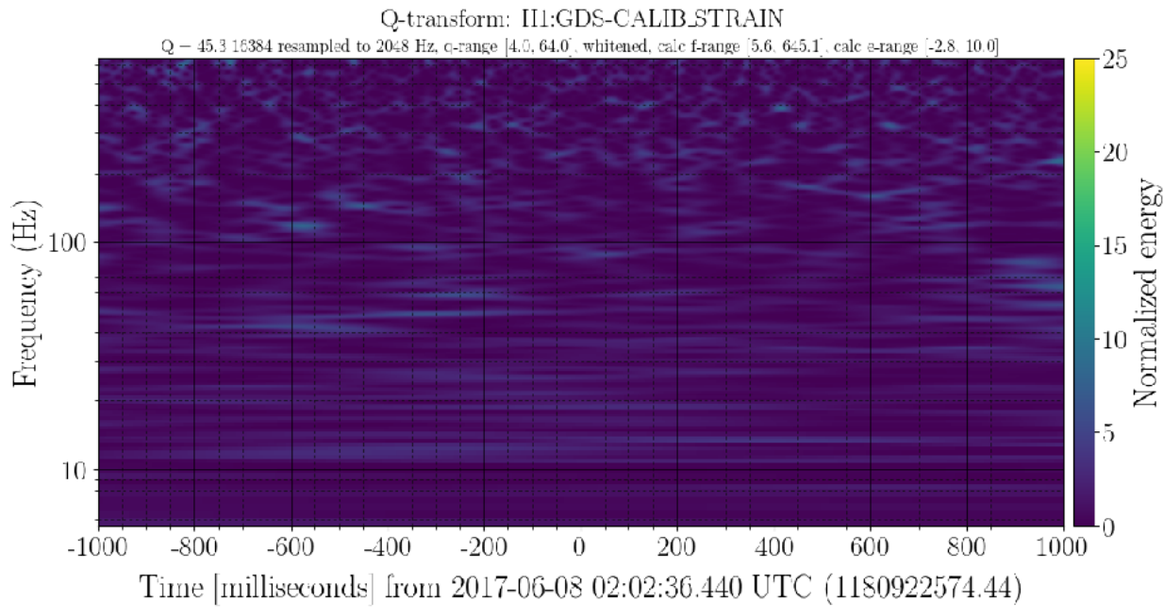
- Data Quality Information
  - Summary page

- Reference

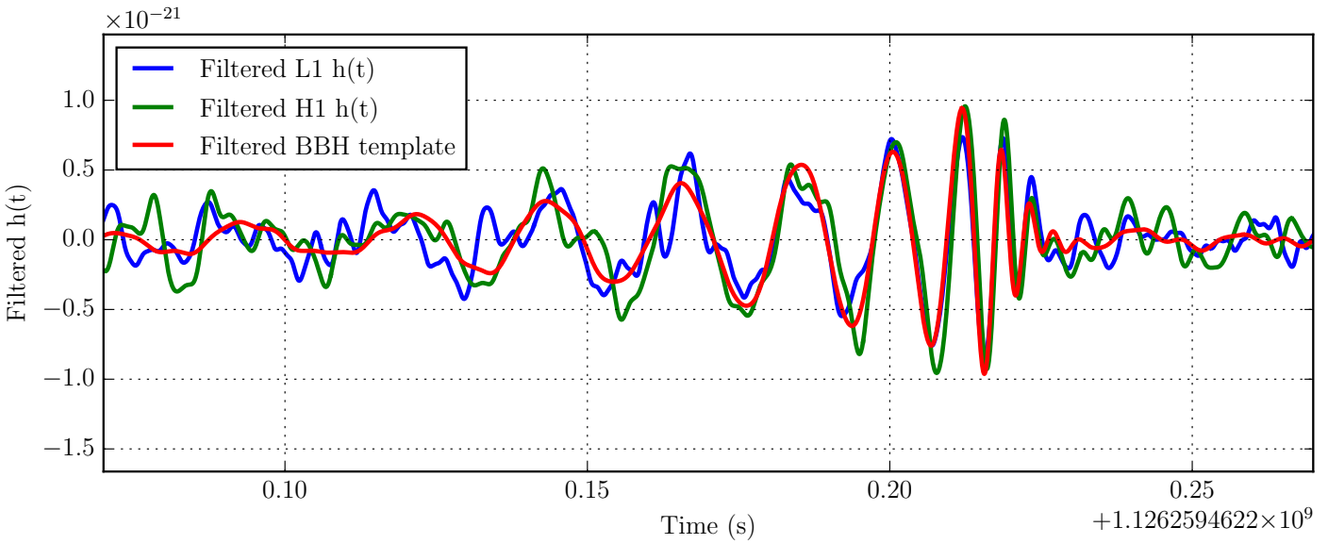
GW data in a perfect world…

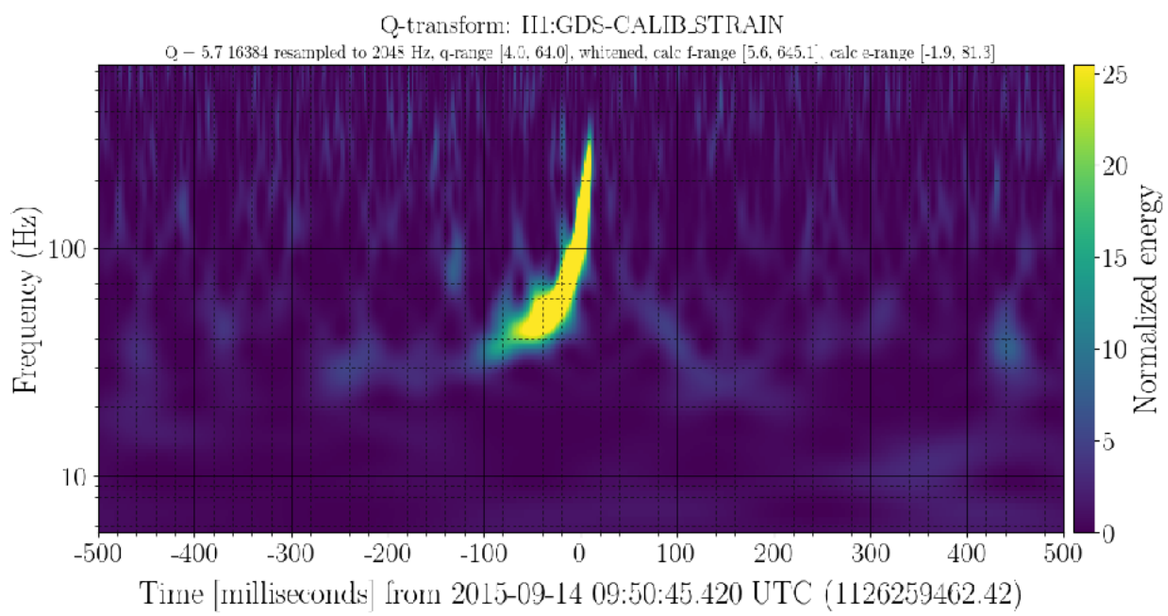
No signal



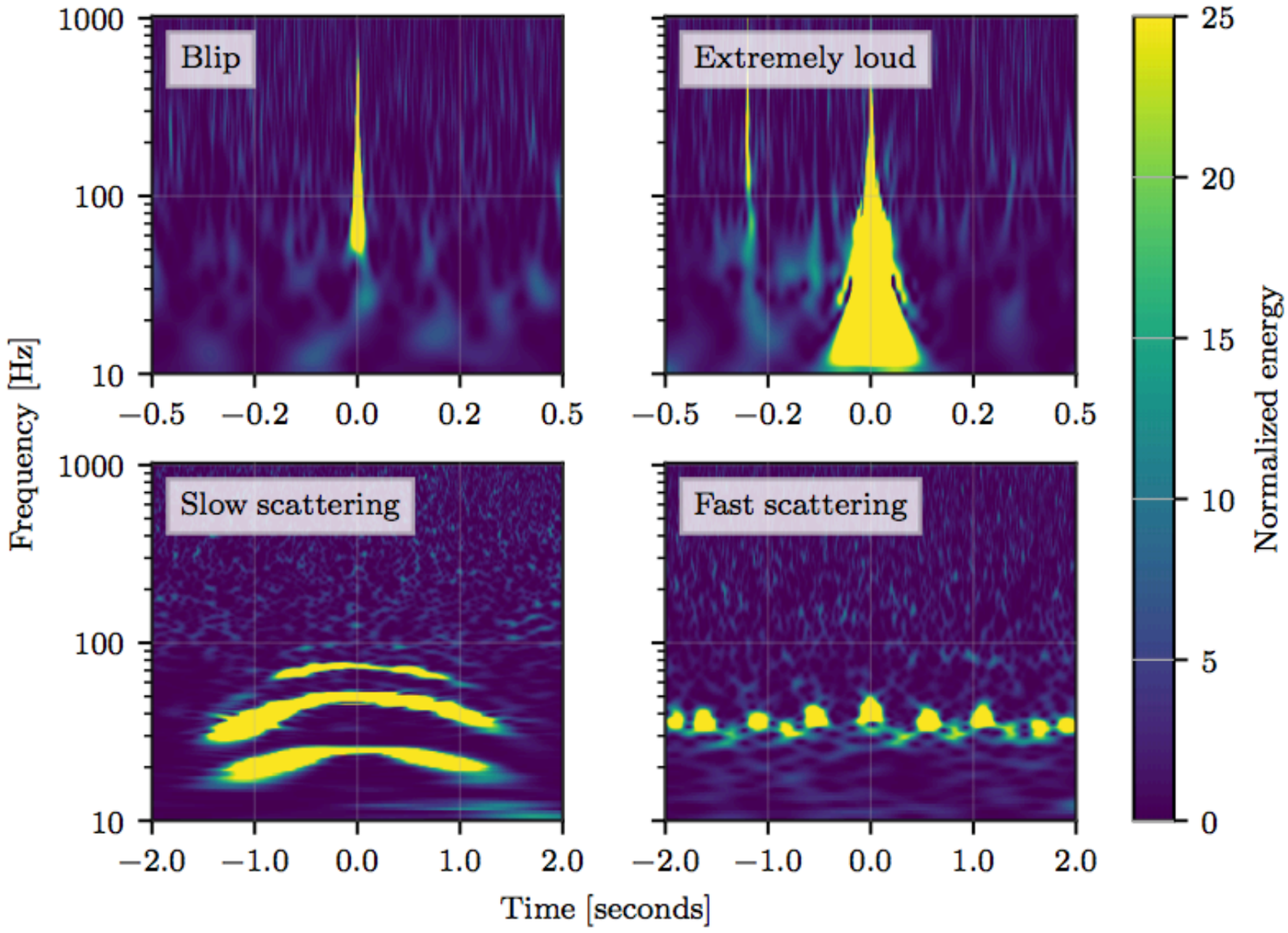


Signal (GW150914)





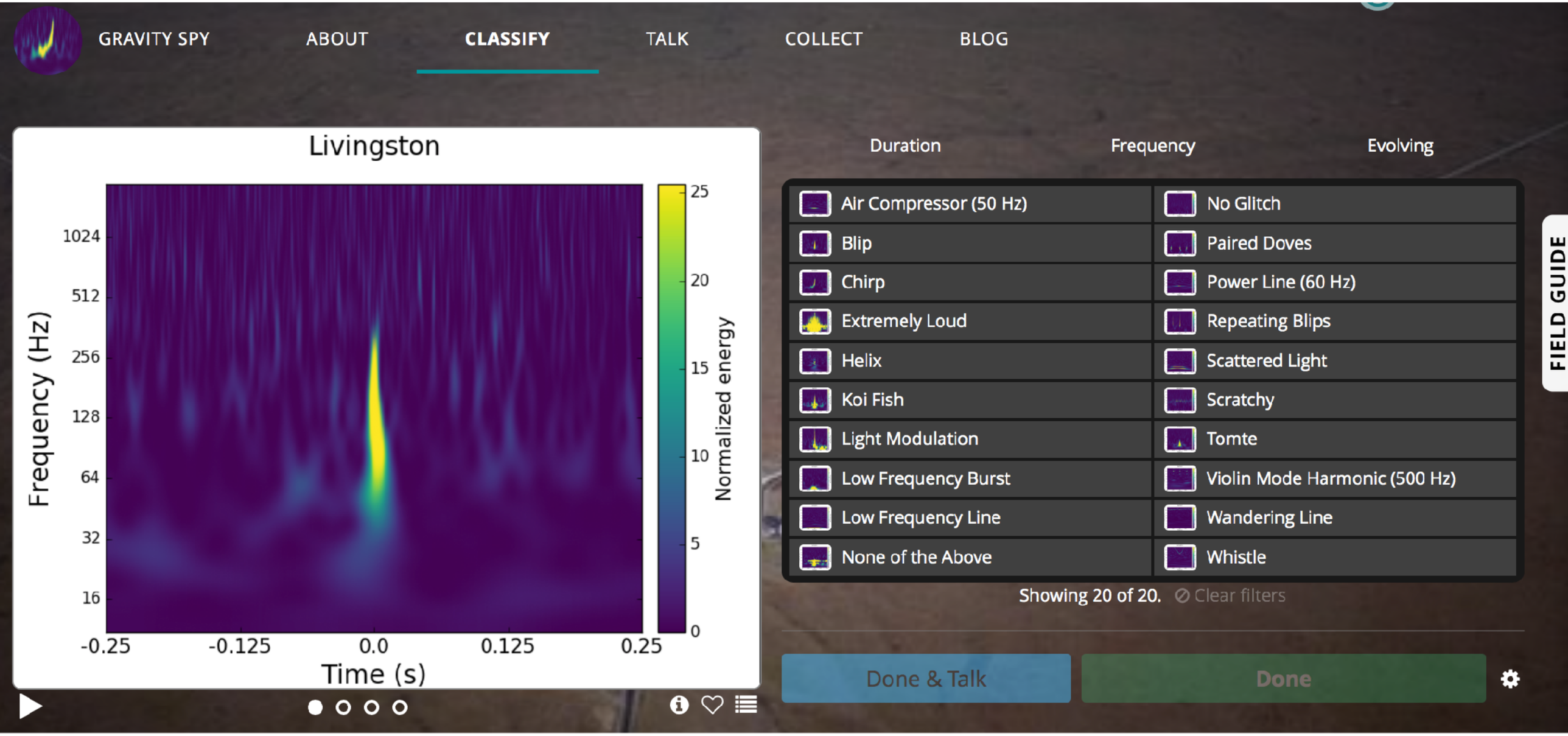
Strain data is non-stationary!



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

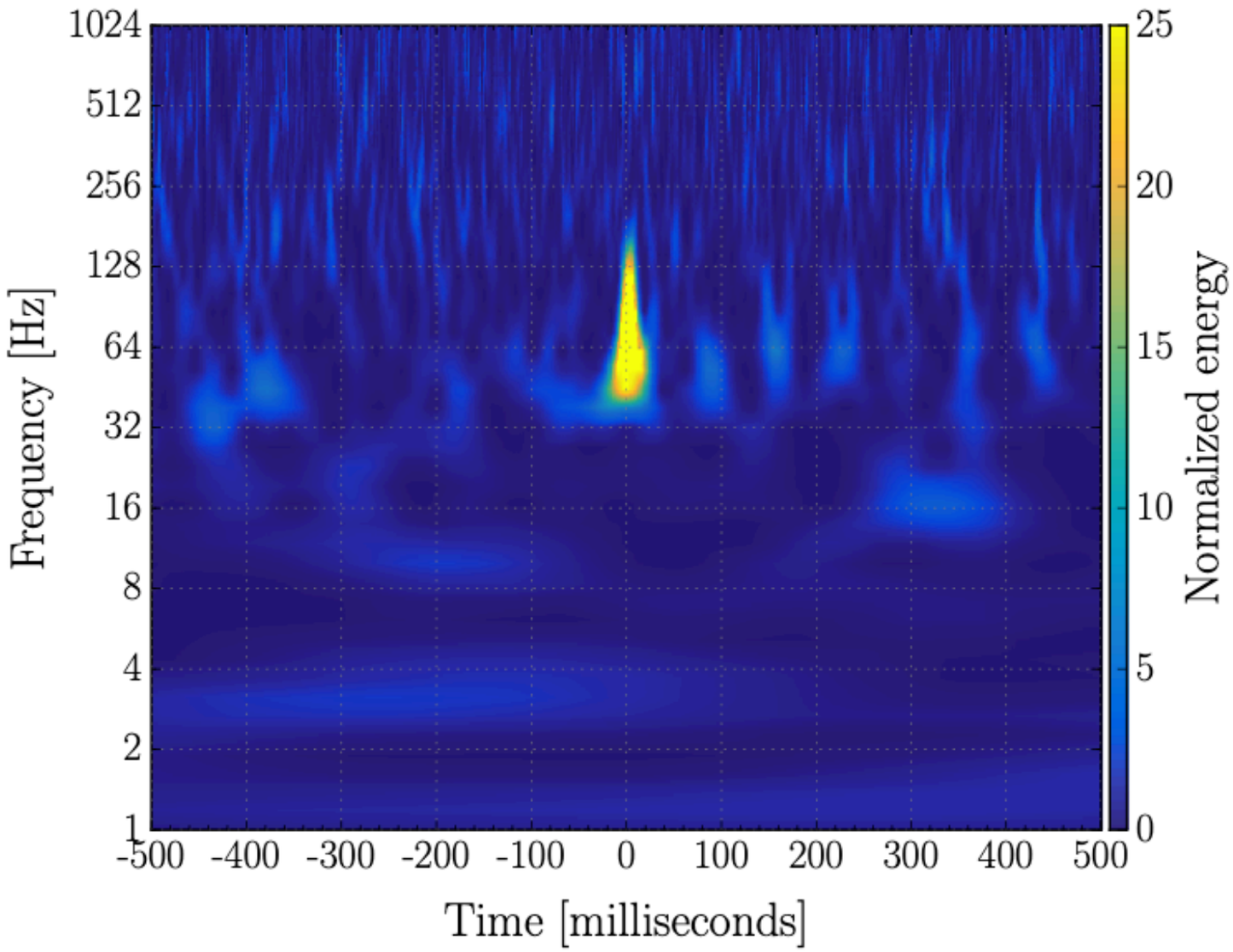
A menagerie of common glitch types

gravityspy.org Zevin et al, 2017, CQG



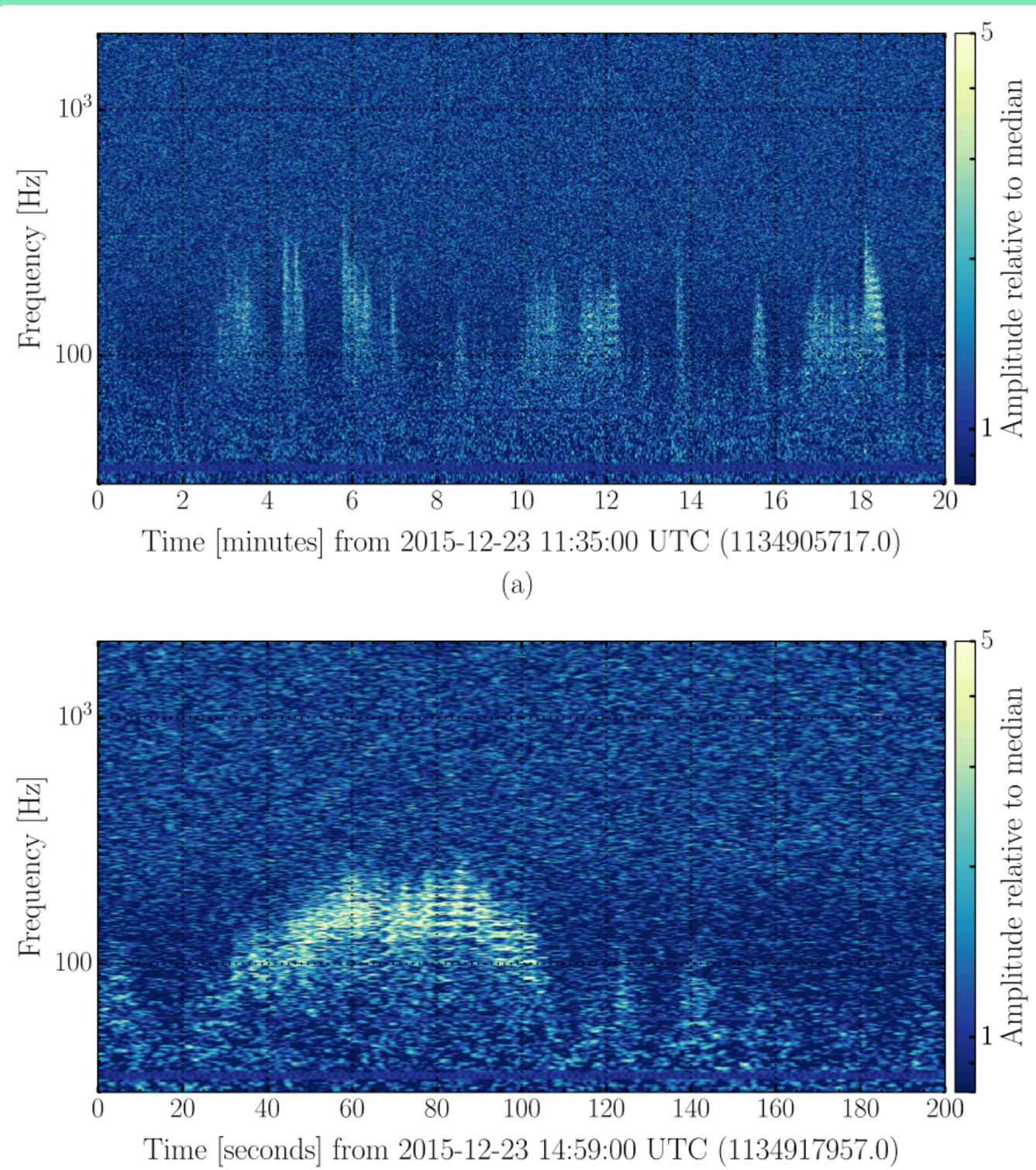
In reality…GW data also contains instrumental and environmental artifacts

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 artifacts

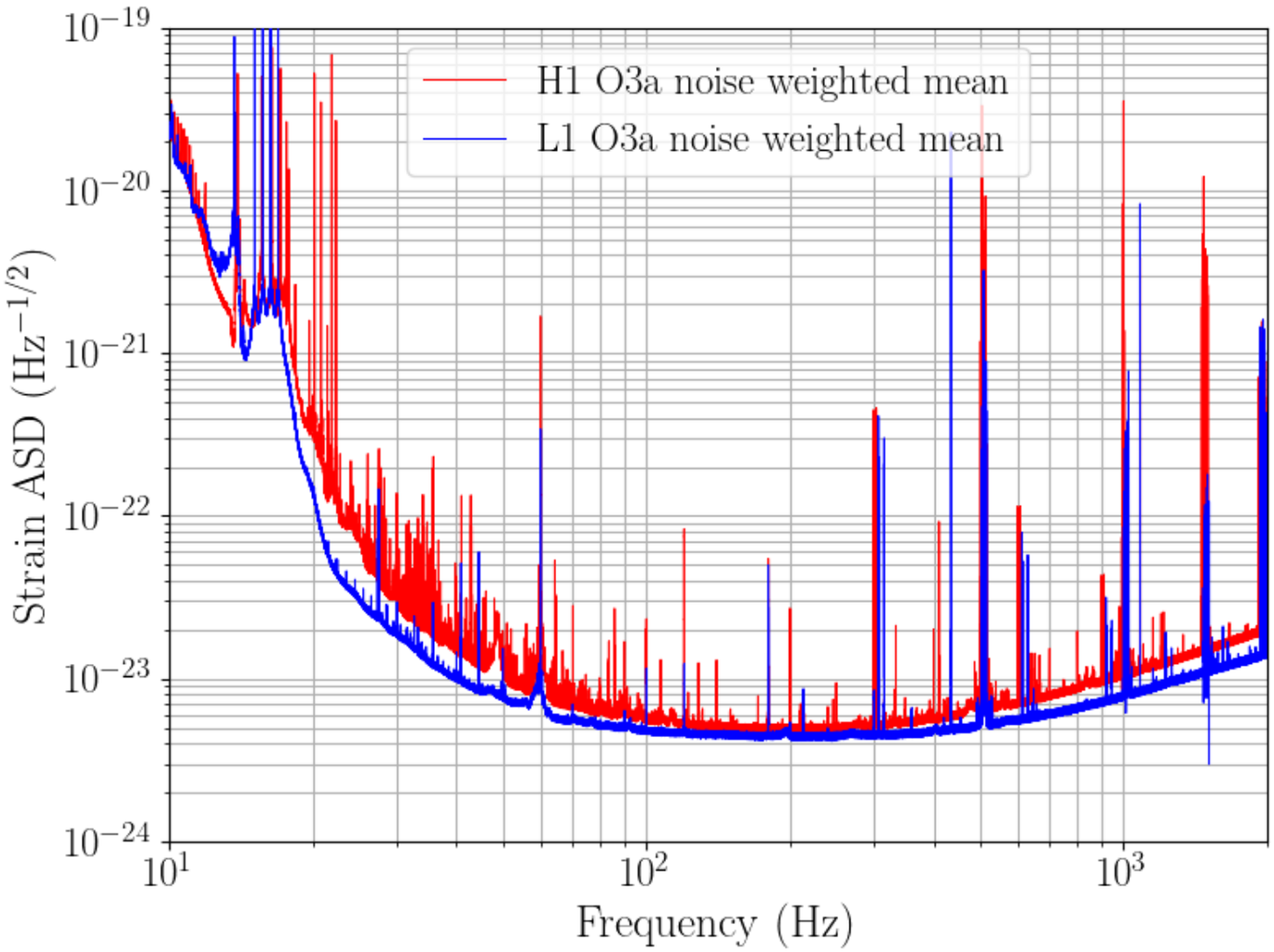


B.P. Abbott et al., *Effects of Data Quality Vetoes on a Search for Compact Binary Coalescences in Advanced LIGO's First Observing Run*, CQG 35, 065010 (2018)

In reality…GW data also contains instrumental and environmental artifacts

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

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



Outline

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

- Data quality: noise artifacts in strain data
  - Glitch
  - Lines

- Mitigating noise artifacts
  - Data quality vetoes
  - Analysis dependent mitigation
  - Event validation

- Data Quality Information
  - Summary page

- Reference

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

See https://gwosc.org/archive/dataset/O3a_16KHZ_R1/

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.

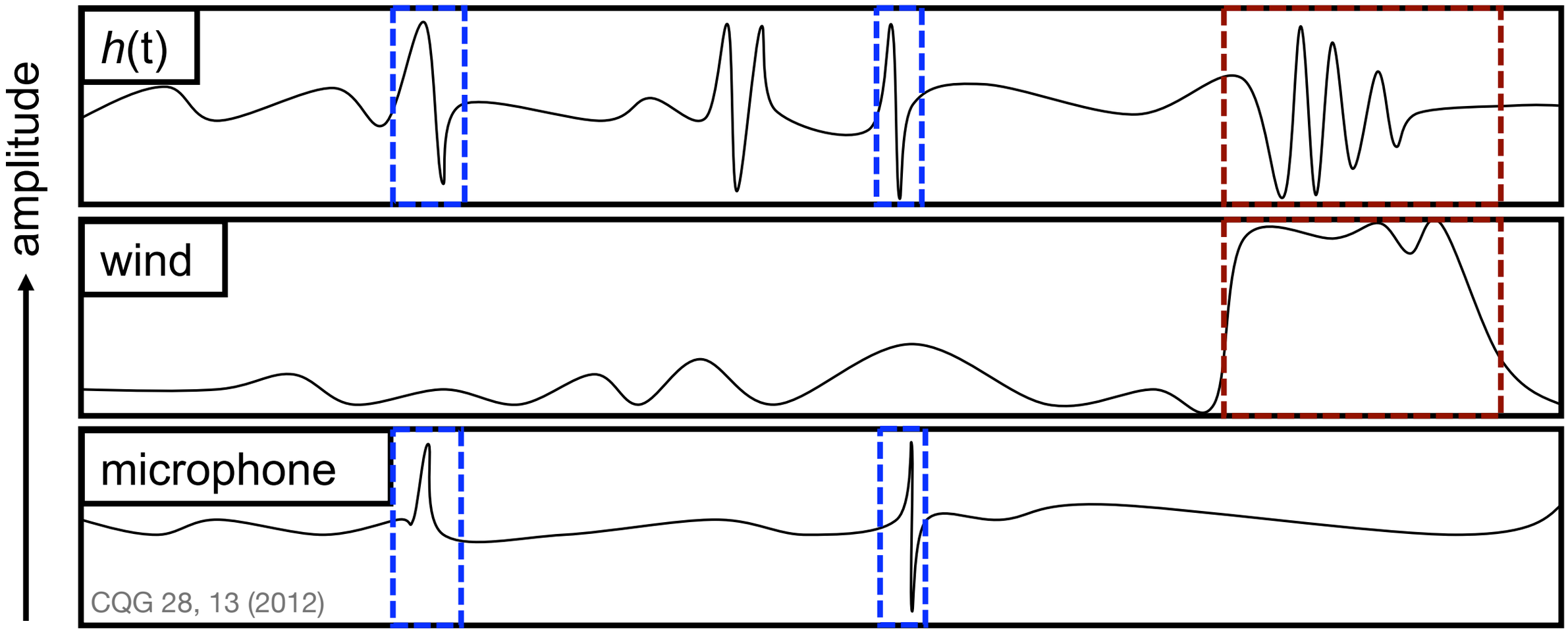
Data quality categories are defined independently for different analysis groups: if something fails at CAT2_BURST, it could pass CAT2_CBC.

Auxiliary channels

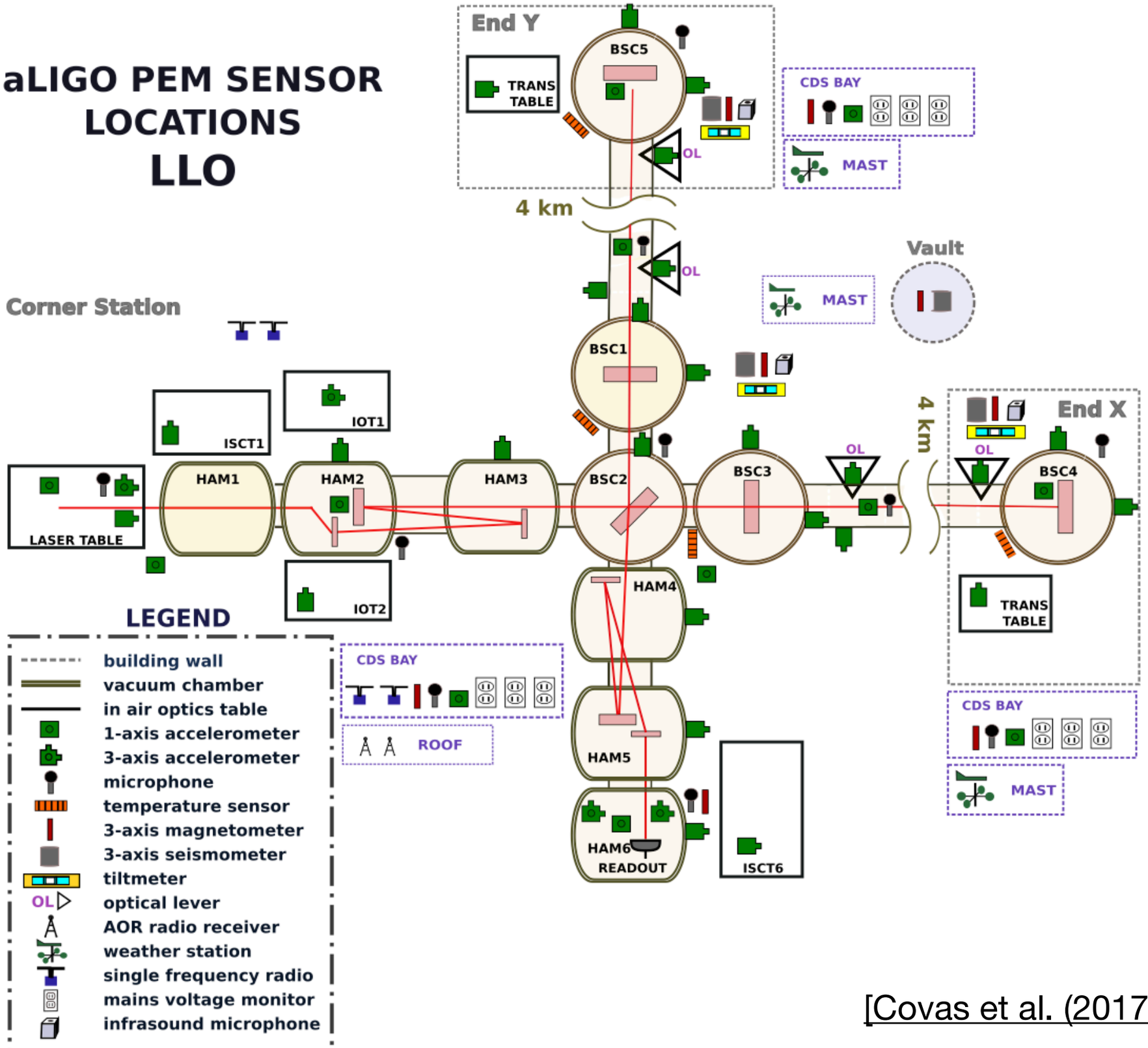
We record **over 200,000 channels per detector** that monitor the environment and detector behavior.

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

Subset of LIGO's auxiliary channels for O3 are publicly available.
See https://gwosc.org/O3/auxiliary/

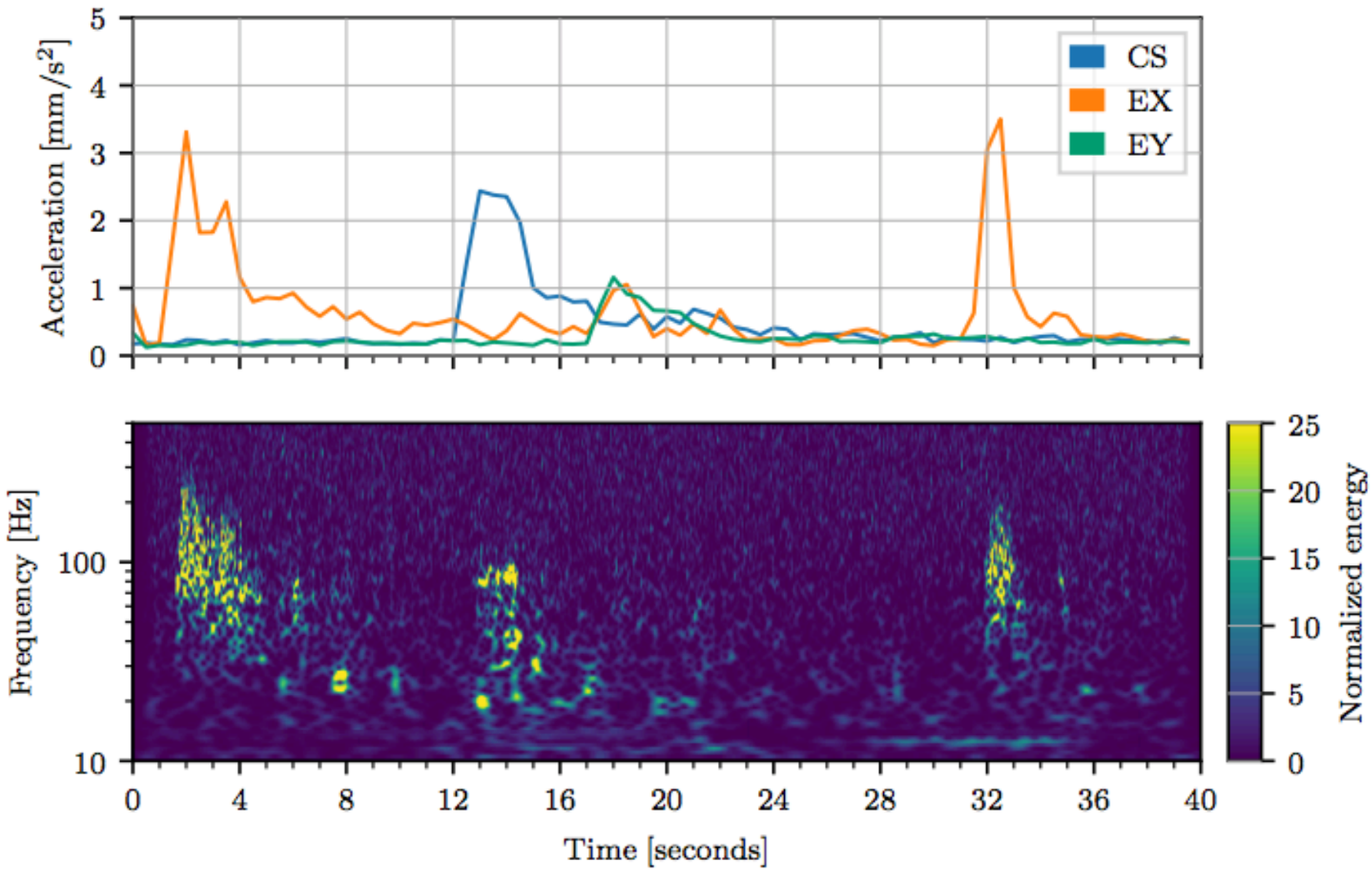


Physical environment channels



[Covas et al. (2017), arXiv 1801.07204]

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)

How to get Data Quality Segments (GWOSC Timeline Query)

Timeline Queries

https://gwosc.org/timeline/query/Run/

The Timeline App shows times when data are available, as well as data quality and injection segments.

Use the **Event Portal** to access individual Events and request any of the Event Timeline or Segment Lists.

Show examples

Select a run
O3a

GPS Start
1238166018

GPS End
1240671618

Duration
2505600

2019-04-01T15:00:00

2019-04-30T15:00:00

Dates shown are in UTC time

Strain Files

Strain Data for V1 | Strain Data for H1 | Strain Data for L1

Segments

Choose the output format below

Plot | JSON | ASCII

Display

☐ H1_BURST_CAT1	☐ L1_BURST_CAT1	☐ V1_BURST_CAT1
☐ H1_BURST_CAT2	☐ L1_BURST_CAT2	☐ V1_BURST_CAT2
☐ H1_BURST_CAT3	☐ L1_BURST_CAT3	☐ V1_BURST_CAT3
☐ H1_CBC_CAT1	☐ L1_CBC_CAT1	☐ V1_CBC_CAT1
☐ H1_CBC_CAT2	☐ L1_CBC_CAT2	☐ V1_CBC_CAT2
☐ H1_CBC_CAT3	☐ L1_CBC_CAT3	☐ V1_CBC_CAT3
☑ H1_DATA	☑ L1_DATA	☑ V1_DATA
☐ H1_NO_BURST_HW_INJ	☐ L1_NO_BURST_HW_INJ	☐ V1_NO_BURST_HW_INJ
☐ H1_NO_CBC_HW_INJ	☐ L1_NO_CBC_HW_INJ	☐ V1_NO_CBC_HW_INJ
☐ H1_NO_CW_HW_INJ	☐ L1_NO_CW_HW_INJ	☐ V1_NO_CW_HW_INJ
☐ H1_NO_DETCHAR_HW_INJ	☐ L1_NO_DETCHAR_HW_INJ	☐ V1_NO_DETCHAR_HW_INJ
☐ H1_NO_STOCH_HW_INJ	☐ L1_NO_STOCH_HW_INJ	☐ V1_NO_STOCH_HW_INJ

Here I selected the O3a data flags for H1, L1, and V1.

How to get Data Quality Segments (GWOSC Timeline Query)

Timeline O3a

From: 2019-04-01T15:00:00 (GPS=1238166018)

To: 2019-04-30T15:00:00 (GPS=1240671618)

Duration: 2505600 s

Strain Data for H1 | Strain Data for L1 | Strain Data for V1

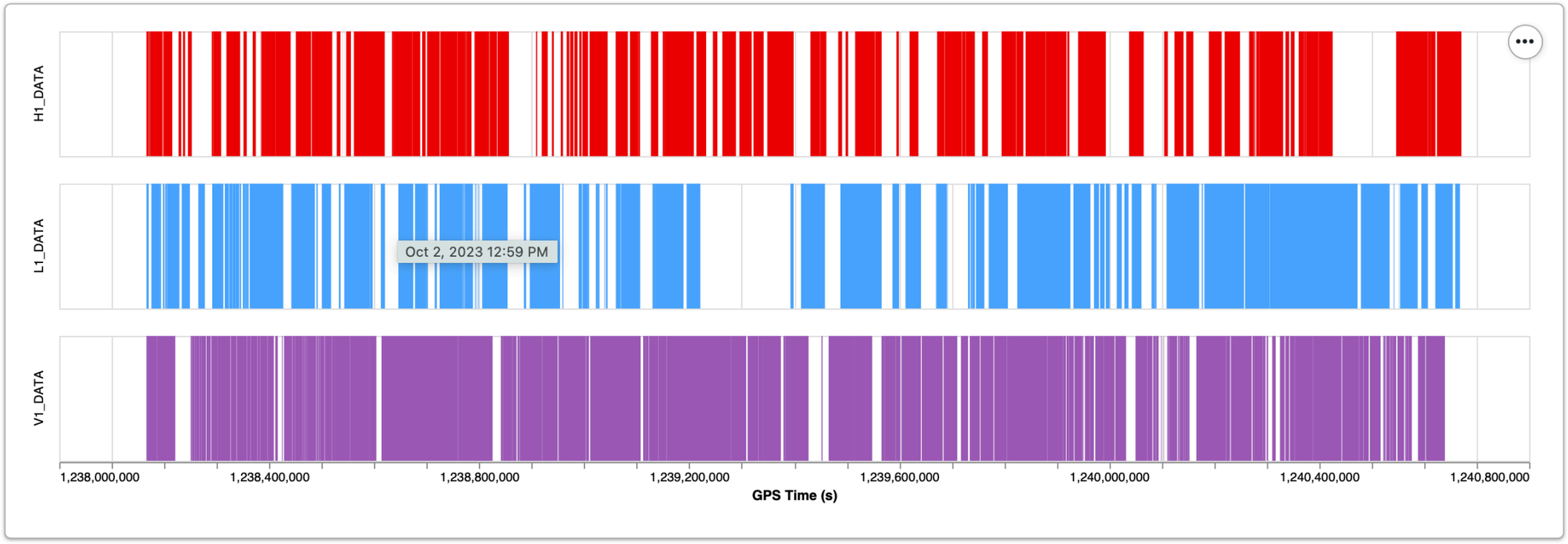
Timeline Stats

	Time Active	Duty Cycle	Segments
H1_DATA	1628079 s	64.98%	107
L1_DATA	1693550 s	67.59%	115
V1_DATA	2196545 s	87.67%	186

Download Segments

JSON | ASCII

H1_DATA | L1_DATA | V1_DATA



https://gwosc.org/timeline/query/Run/

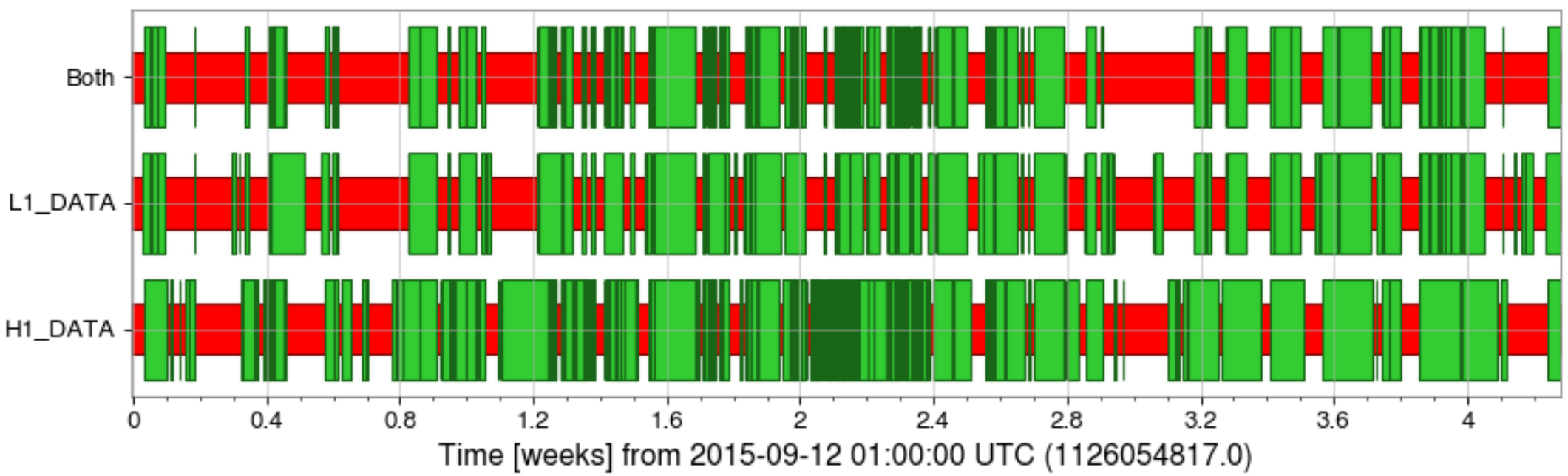
Segments can be plotted (with interactive zooming) or downloaded

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/

```
l1month1 = DataQualityFlag.fetch_open_data('L1_DATA', 'Sep 12 2015',
                                           'Oct 12 2015')
bothon = h1month1 & l1month1
plot = h1month1.plot()
ax = plot.gca()
ax.plot(l1month1)
ax.plot(bothon, label='Both')
plot.show()
```



Outline

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

- Data quality: noise artifacts in strain data
  - Glitch
  - Lines

- Mitigating noise artifacts
  - Data quality vetoes
  - Analysis dependent mitigation
  - Event validation

- Data Quality Information
  - Summary page

- Reference

Daily detector status (available for O2+O3+O4a)

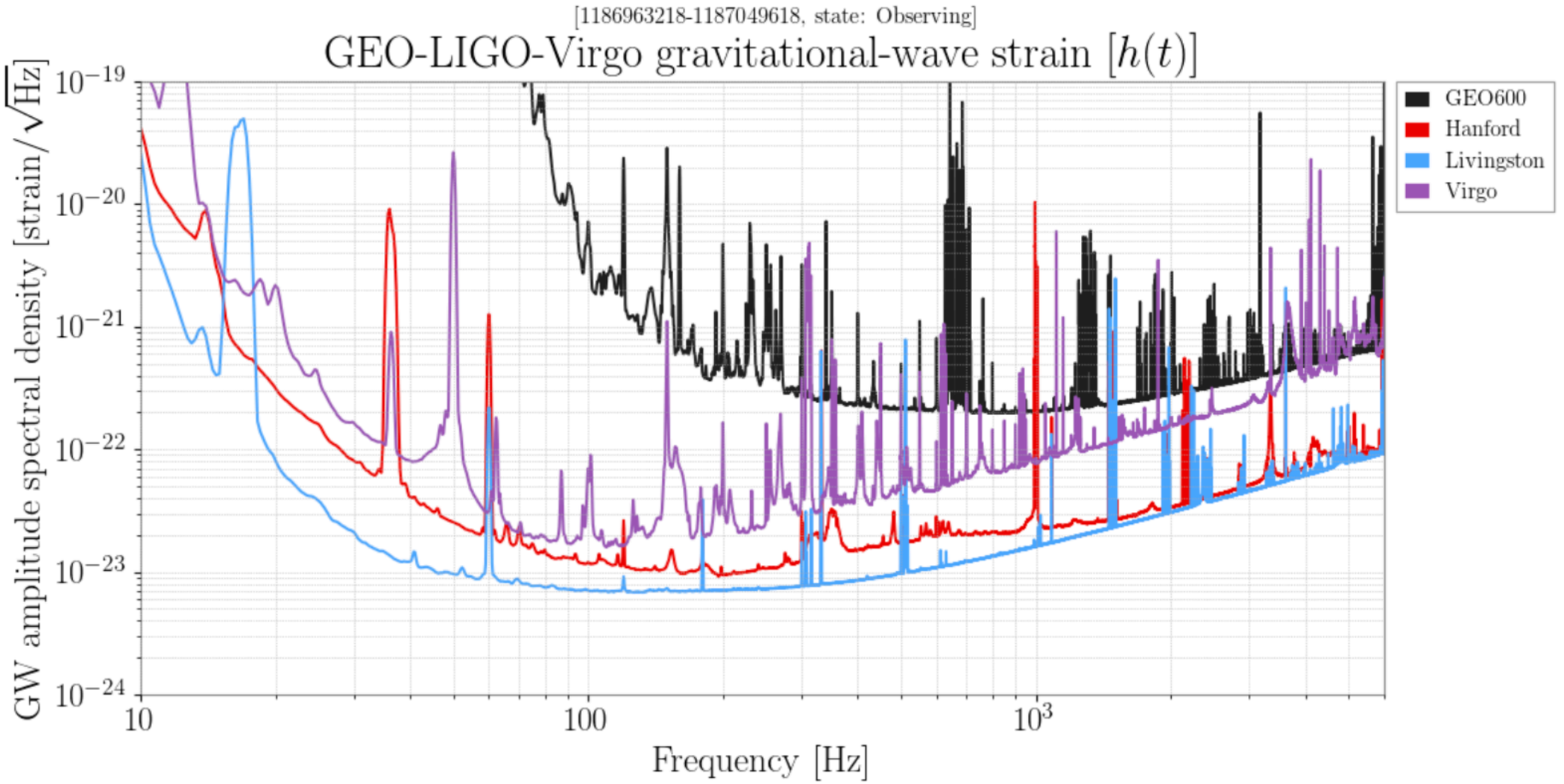
https://www.gw-openscience.org/detector_status/day/20170817/

« August 17 2017 » Summary Home Environment Instrument performance

Date selection

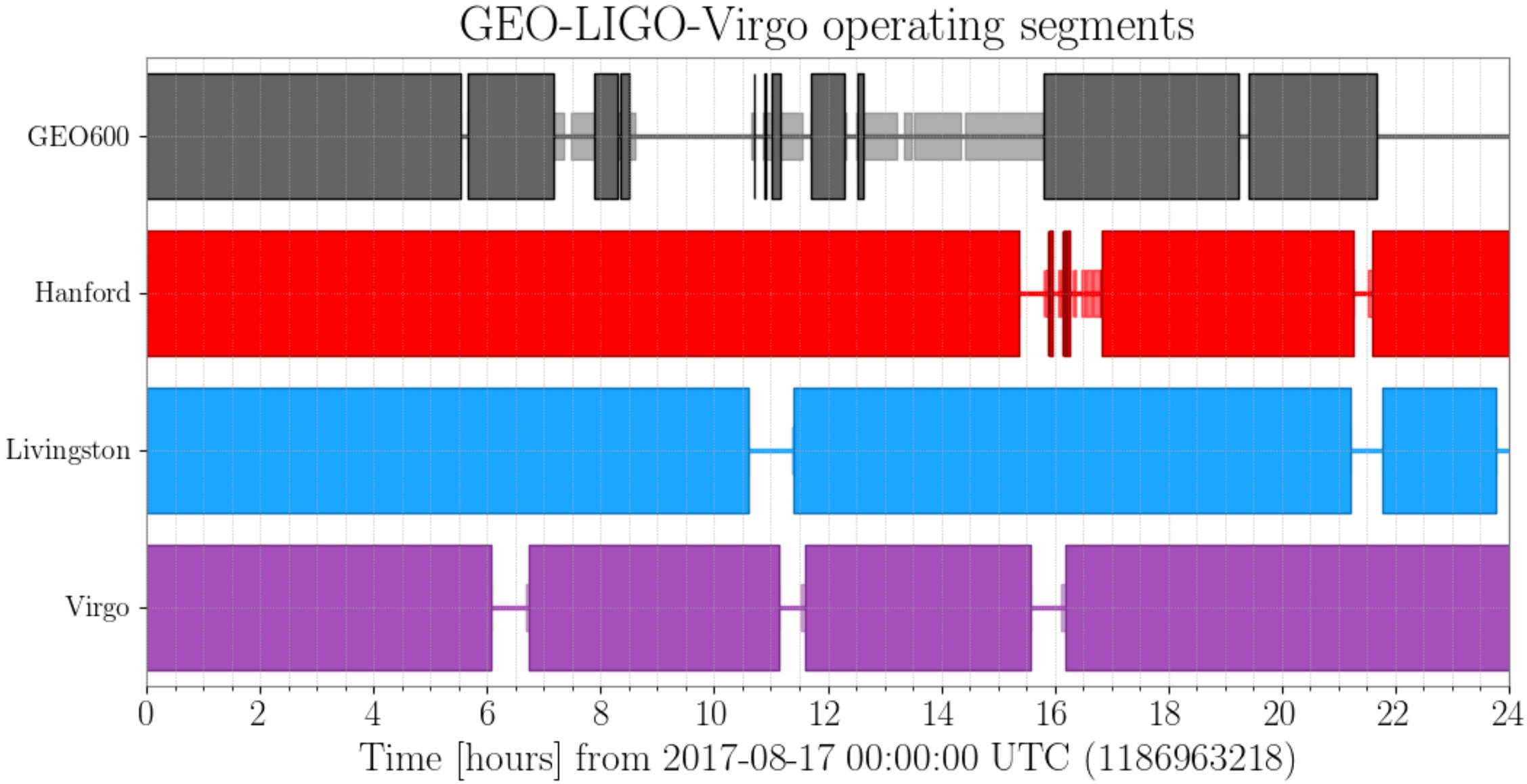
Summary

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.



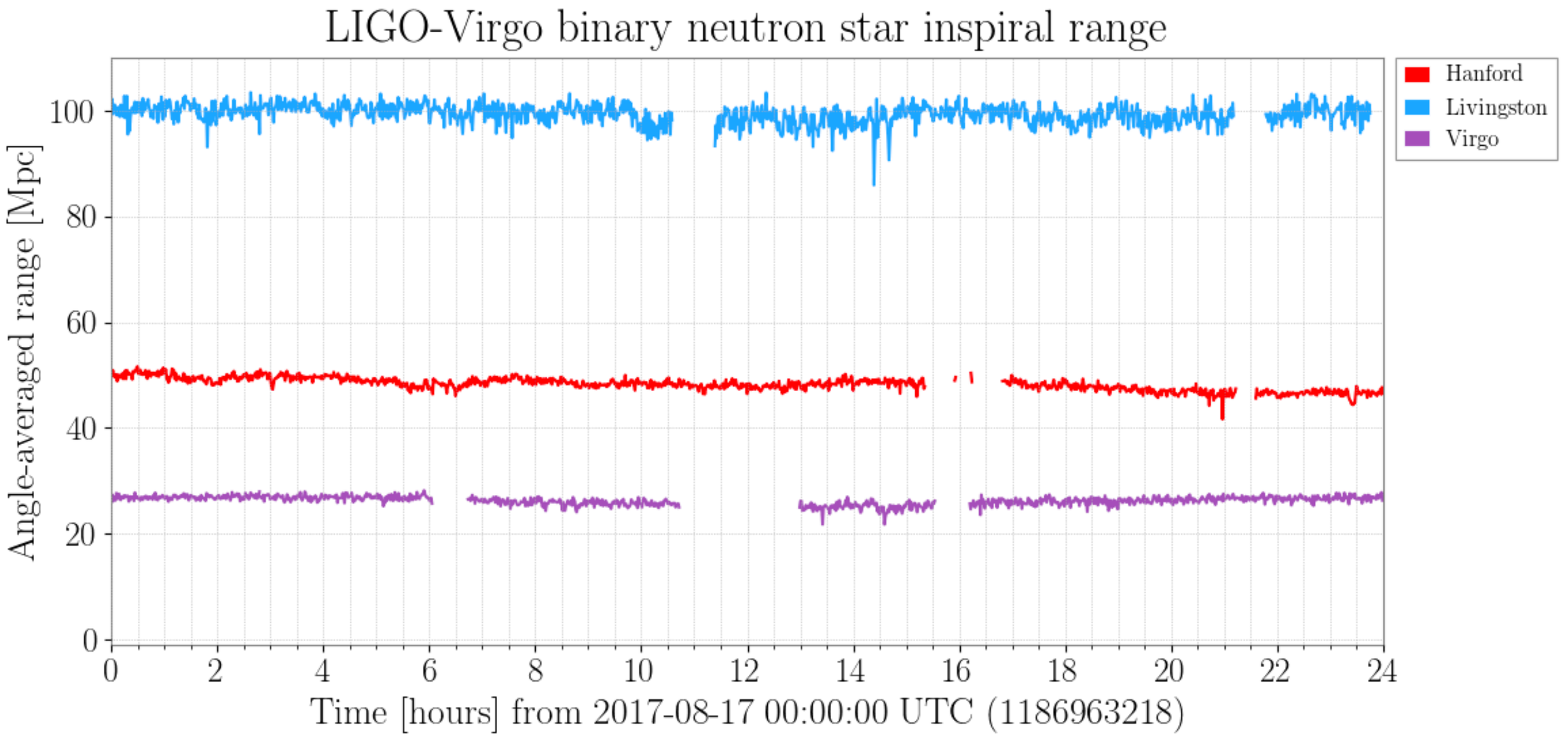
Daily detector status (available for O2+O3+O4a)

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



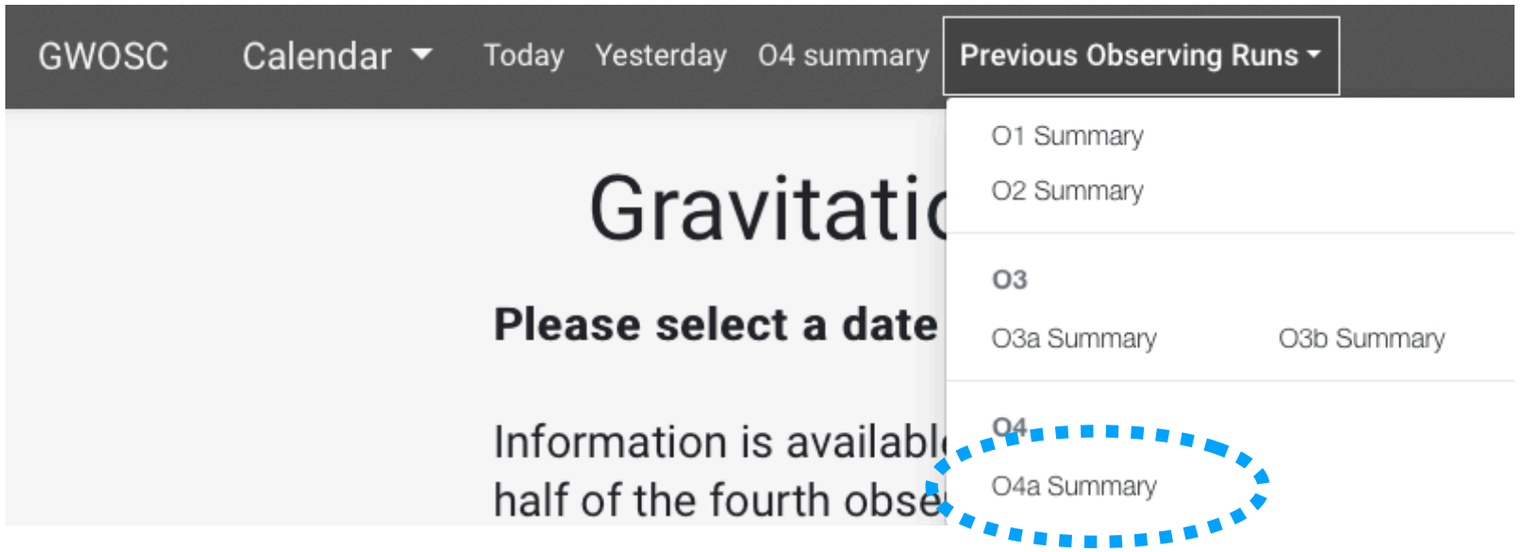
Daily detector status (available for O2+O3+O4a)

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



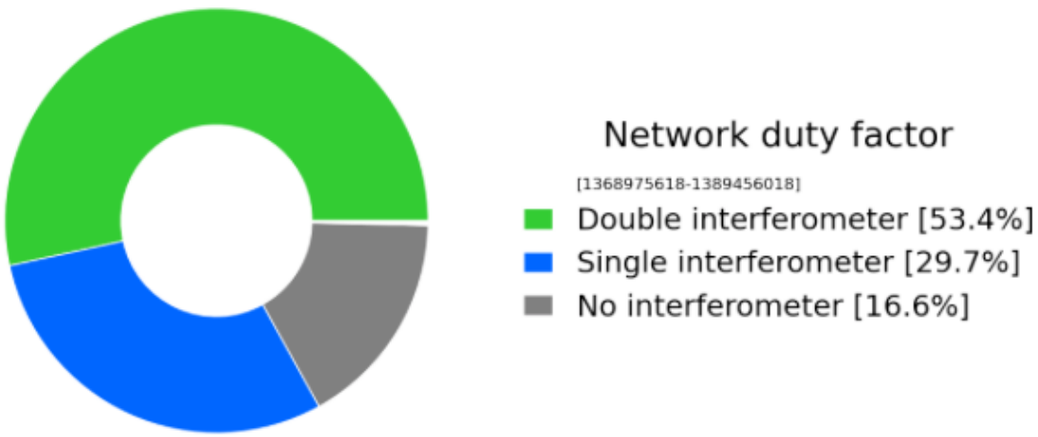
Observing run summaries (O1+O2+O3+O4a)

https://gwosc.org/detector_status/



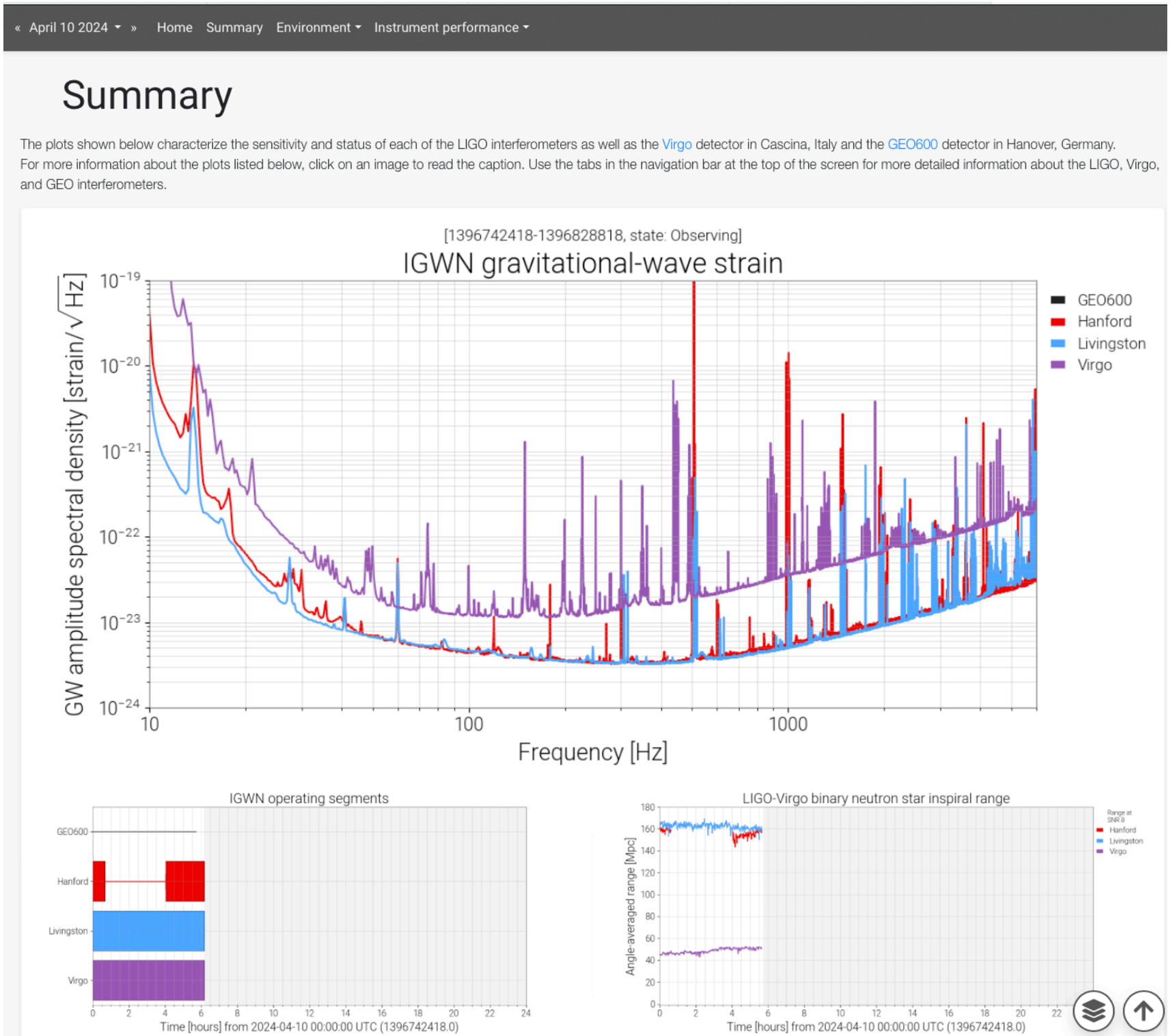
https://gwosc.org/detector_status/O4a/

Includes summary plots of LIGO segments and sensitivity over the run



Observing run summaries (O4b)

https://gwosc.org/detector_status/day/20240410/



Outline

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

- Data quality: noise artifacts in strain data
  - Glitch
  - Lines

- Mitigating noise artifacts
  - Data quality vetoes
  - Analysis dependent mitigation
  - Event validation

- Data Quality Information
  - Summary page

- Reference

Useful data quality references

For glitches:
GW150914 Detector Characterization paper: arXiv 1602.03844
O2/O3 LIGO Detector Chracterization paper: arXiv: 2101.11673
O3 Virgo paper: arXiv: 2205.01555
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/
O3 lines calico on GWOSC: https://www.gw-openscience.org/O3/o3aspeclines/

Data Quality around events:
GWTC-2 paper: arXiv: 2010.14527
GWTC-3 paper: arXiv: 2111.03606

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/

O4a public alerts: https://gracedb.ligo.org/superevents/public/O4/

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