LIGO strain data and data quality

Made with GWpy: gwpy.github.io

Dr. Jess McIver

LIGO Open Data Workshop
March 25, 2018

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

What's in a LIGO data file?

meta: Meta-data for the file. This is **basic information** such as the GPS times covered, which instrument, etc.

strain: Strain data from the interferometer. This is "the data", the **main measurement of spacetime strain** recorded by the LIGO detectors.

quality: A 1 Hz time series describing the **data quality** for each second of data.

LIGO Open Science Center Tutorial #2
https://losc.ligo.org/tutorial02/

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

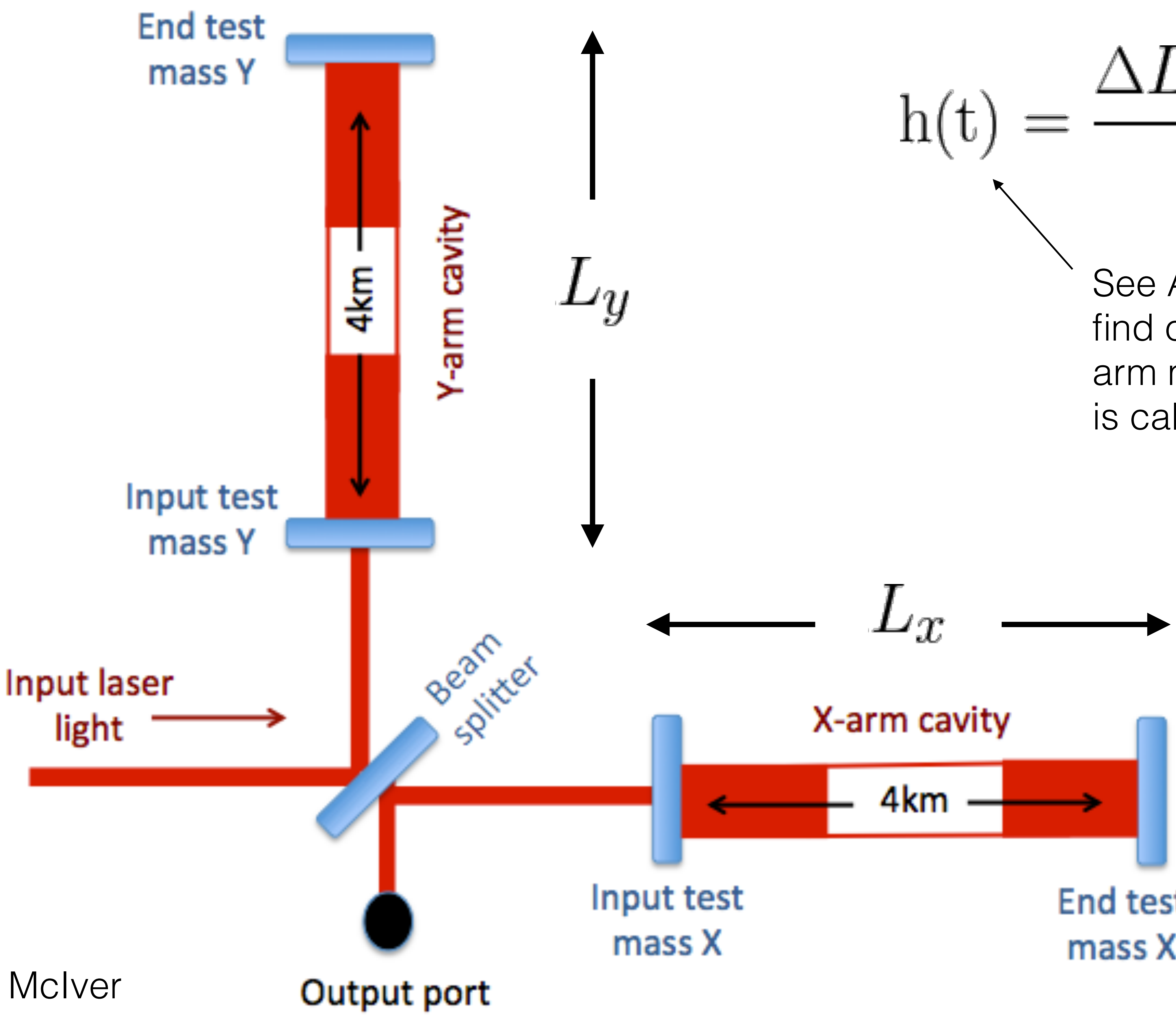
- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

What is LIGO strain data, h(t) ?

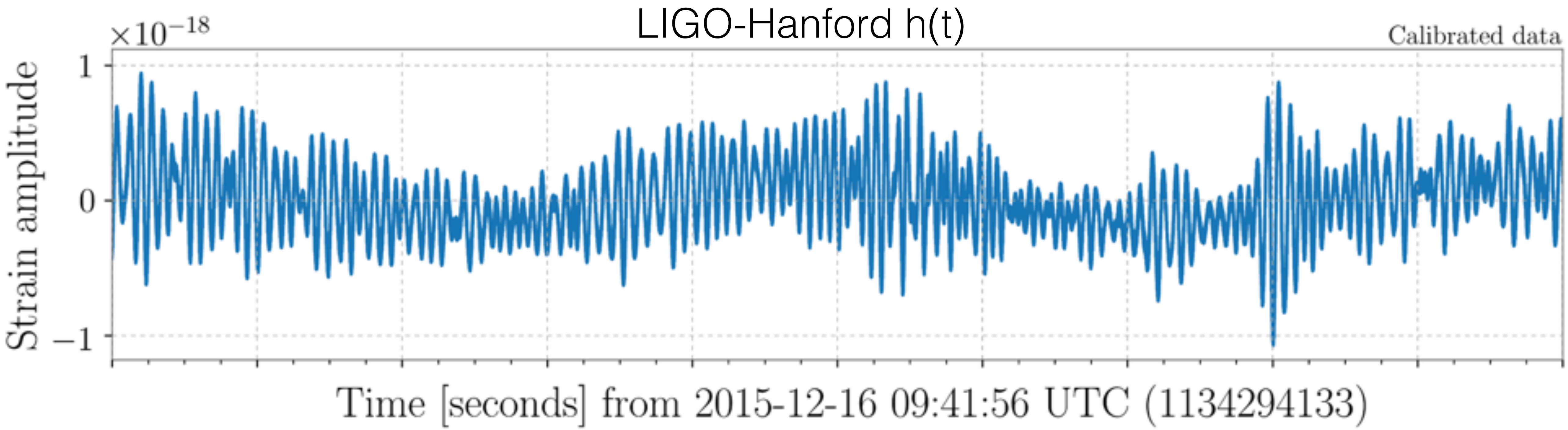


$$\text{h(t)} = \frac{\Delta L_y - \Delta L_x}{L}$$

See Alex Urban's talk to find out how the differential arm motion control signal is calibrated into h(t)

McIver

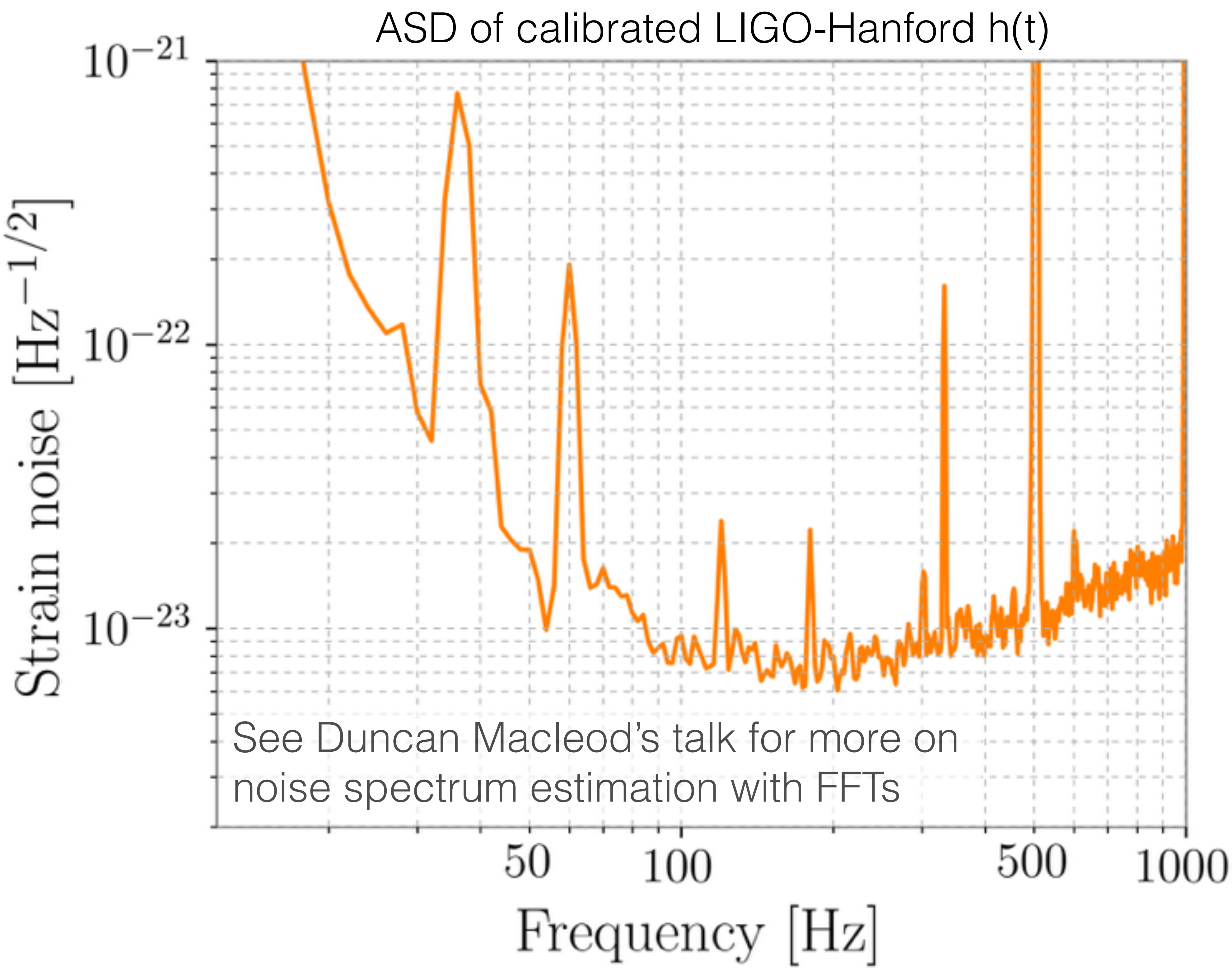
What does LIGO data look like?



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

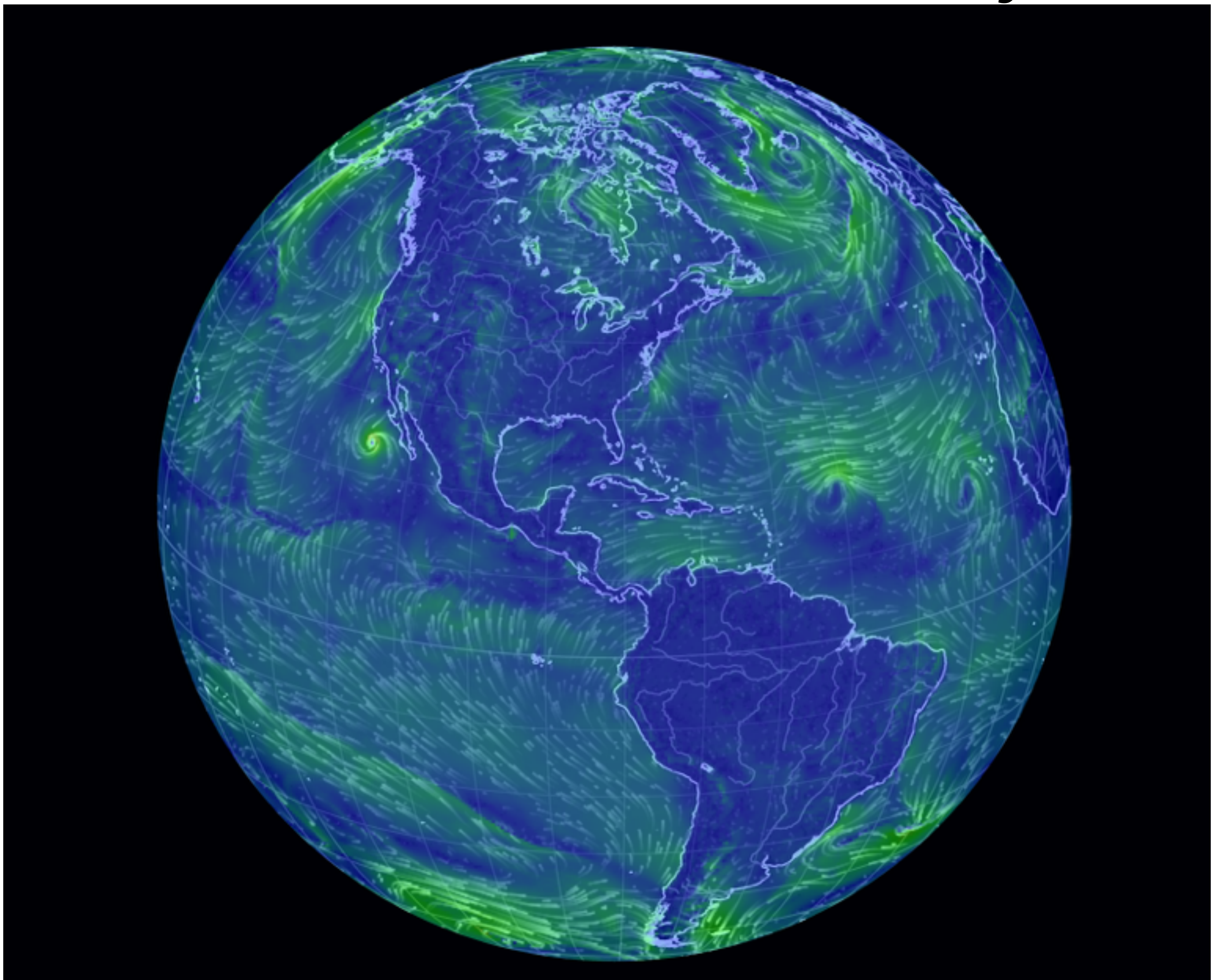
Made with GWpy by Duncan Macleod

LIGO data in the frequency domain



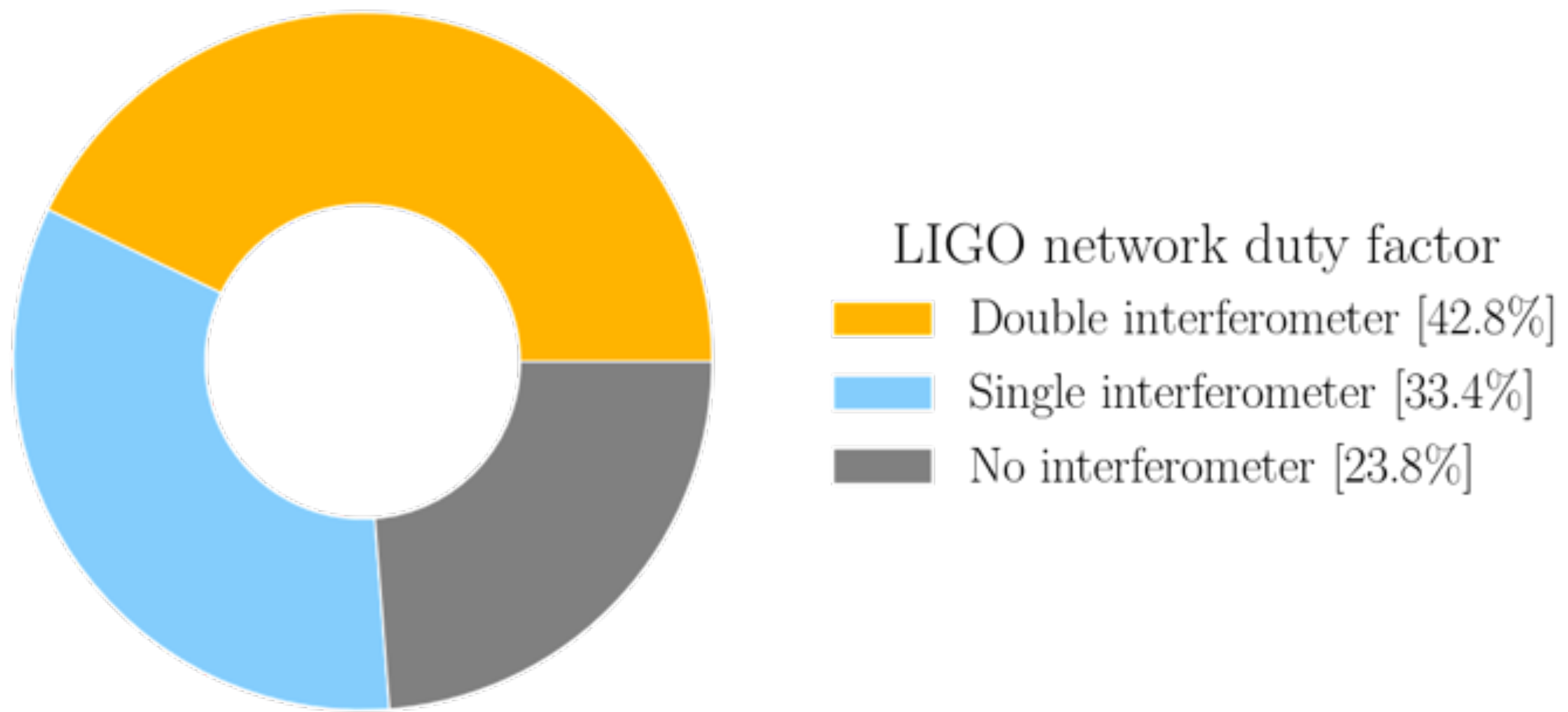
Made with GWpy by Duncan Macleod

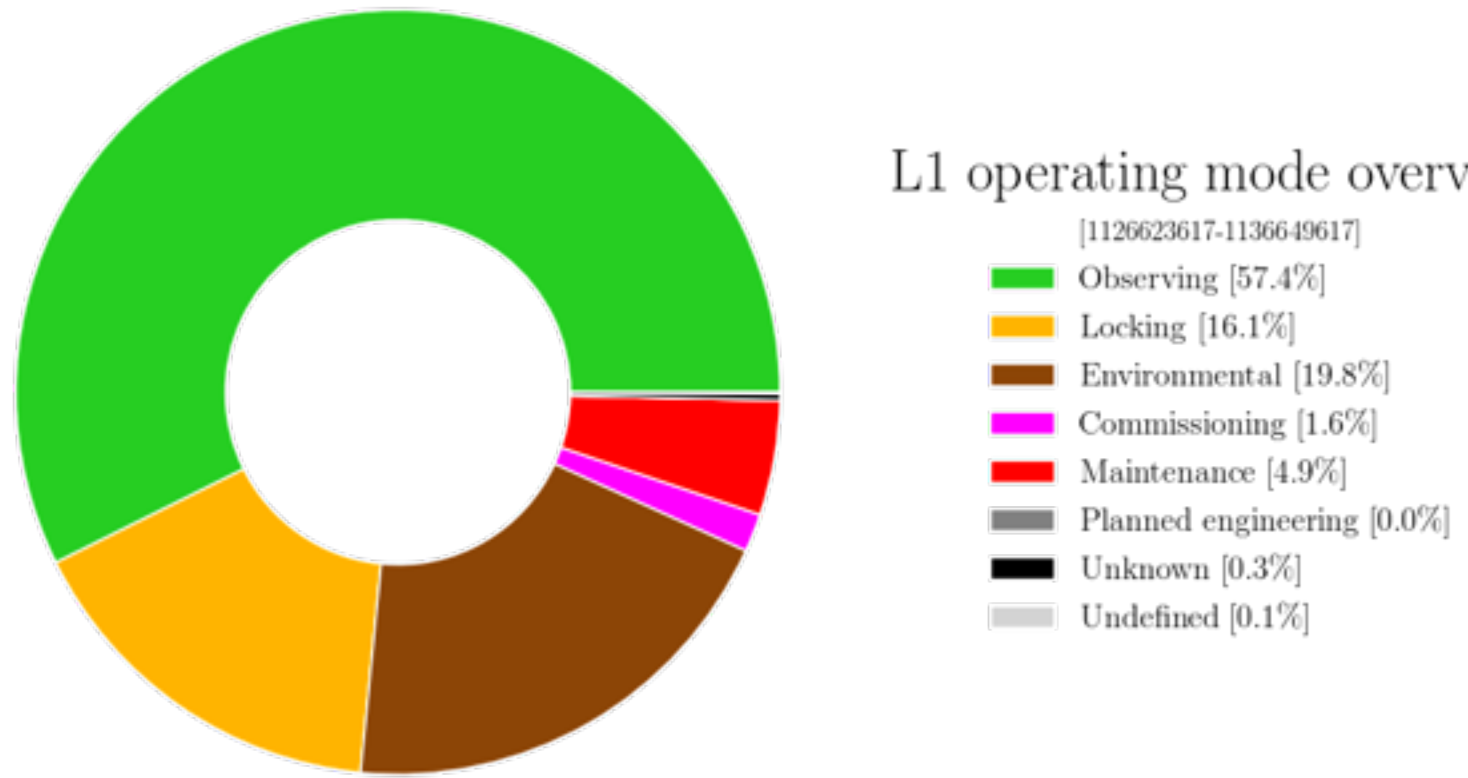
Interferometer stability



earth.nullschool.net

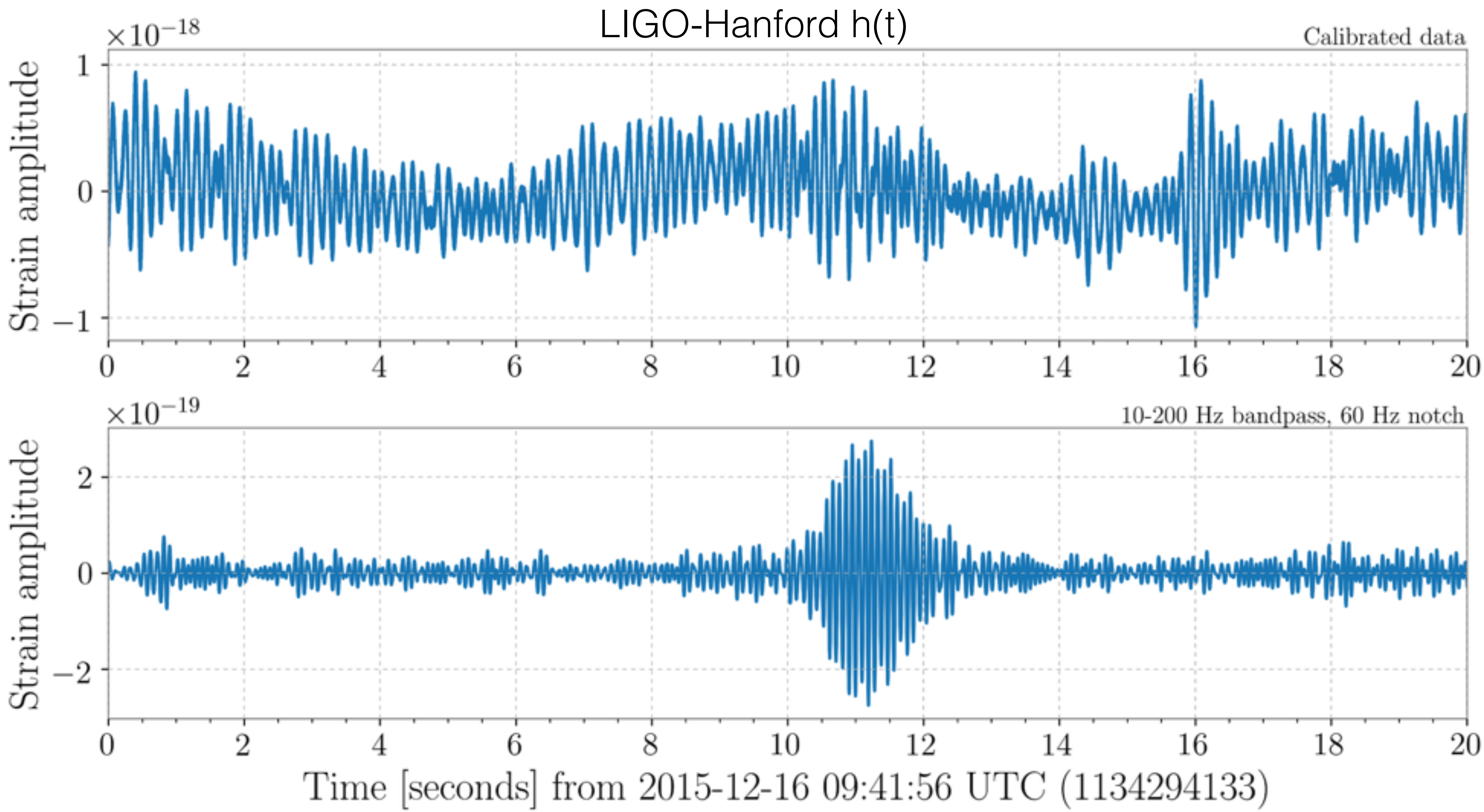
O1 network interferometer uptime





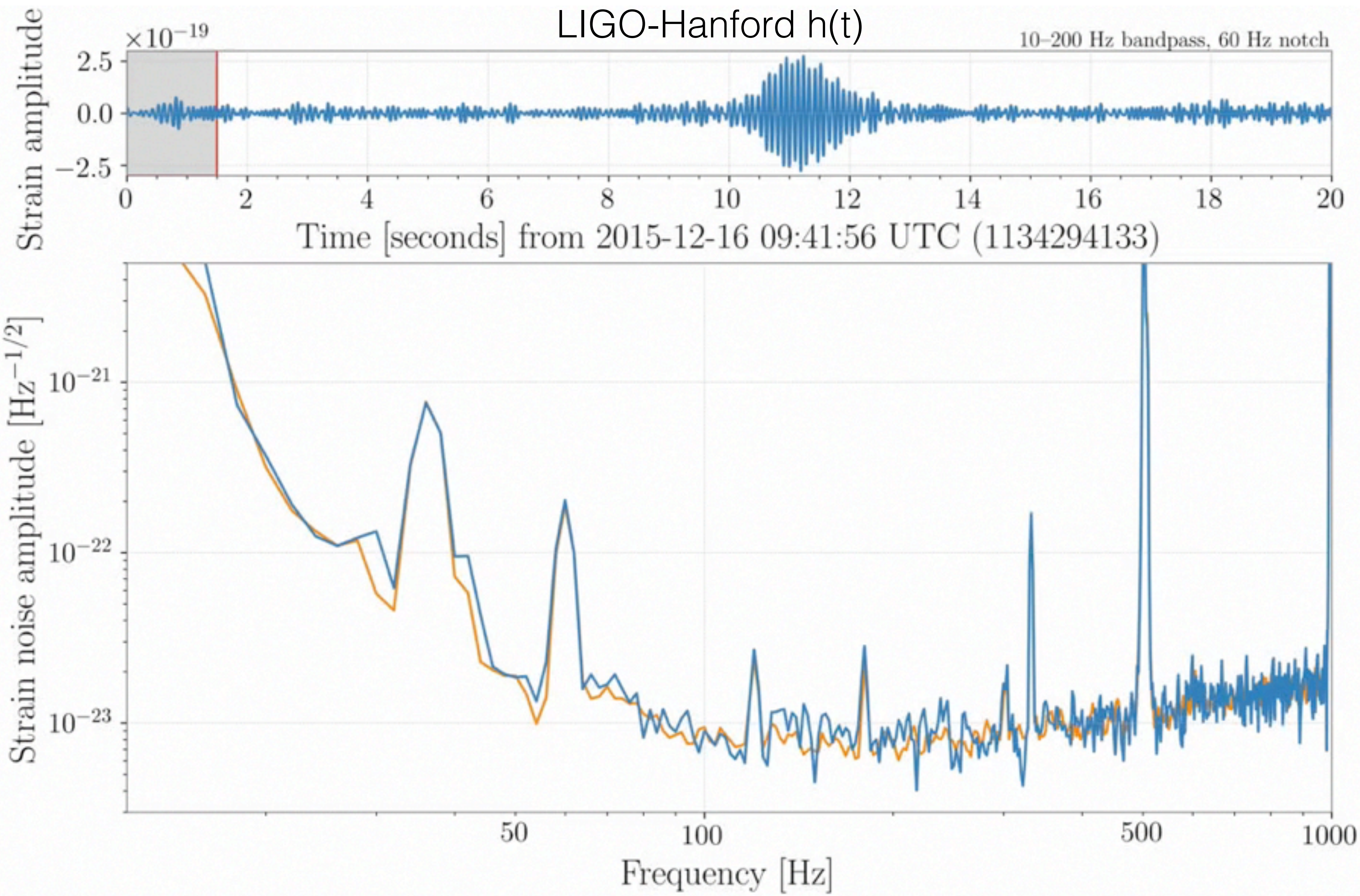
This is a rough estimate based on human input rather than an automated state vector

What does LIGO data look like?



Made with GWpy by Duncan Macleod

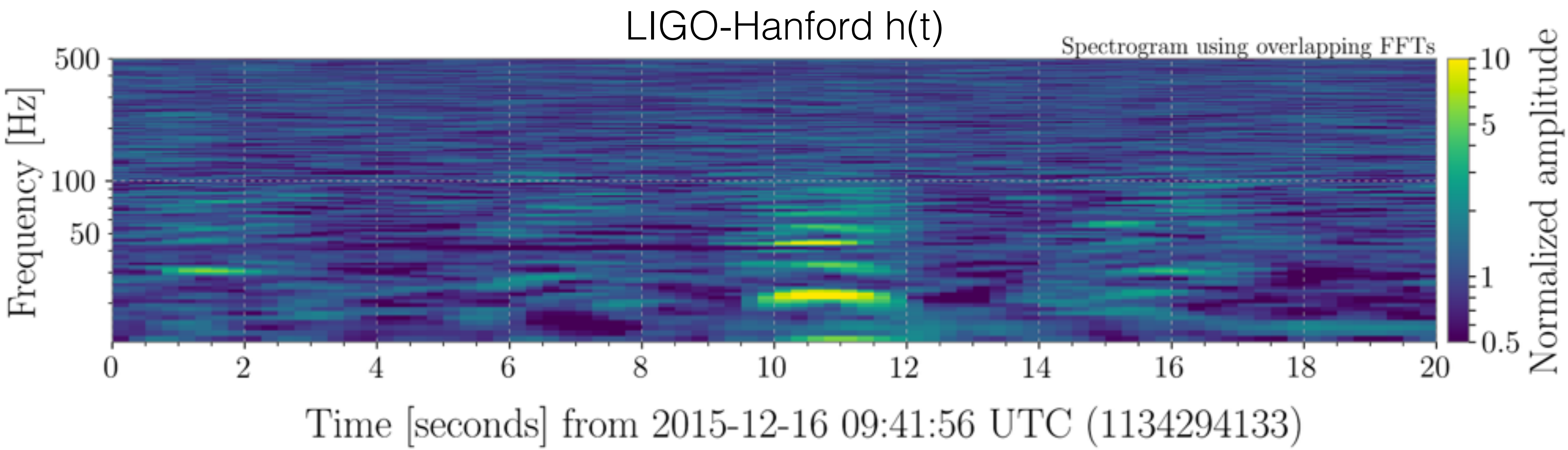
LIGO data in the frequency domain



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

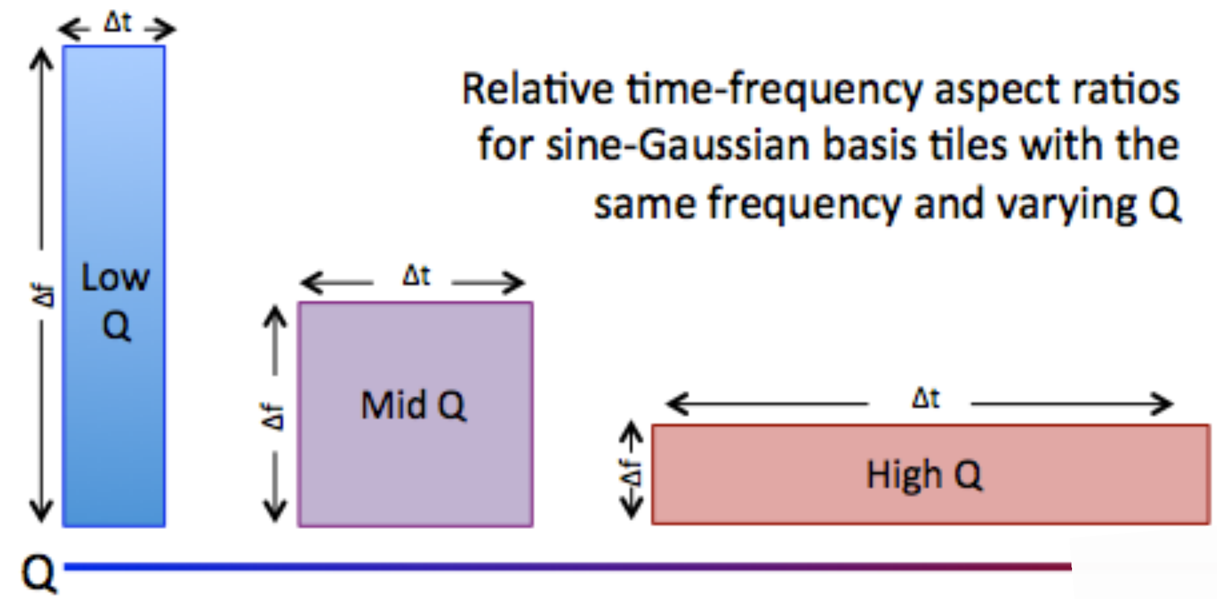
Time-frequency spectrogram



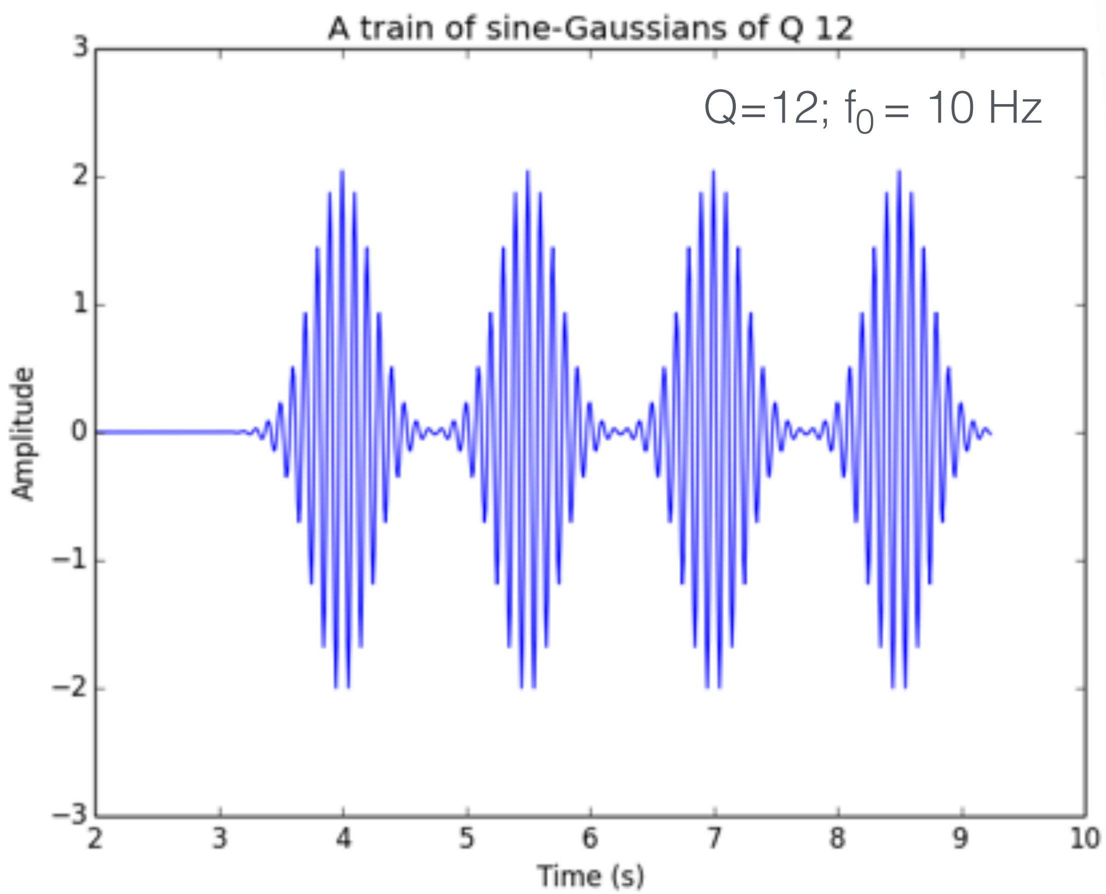
Made with GWpy by Duncan Macleod

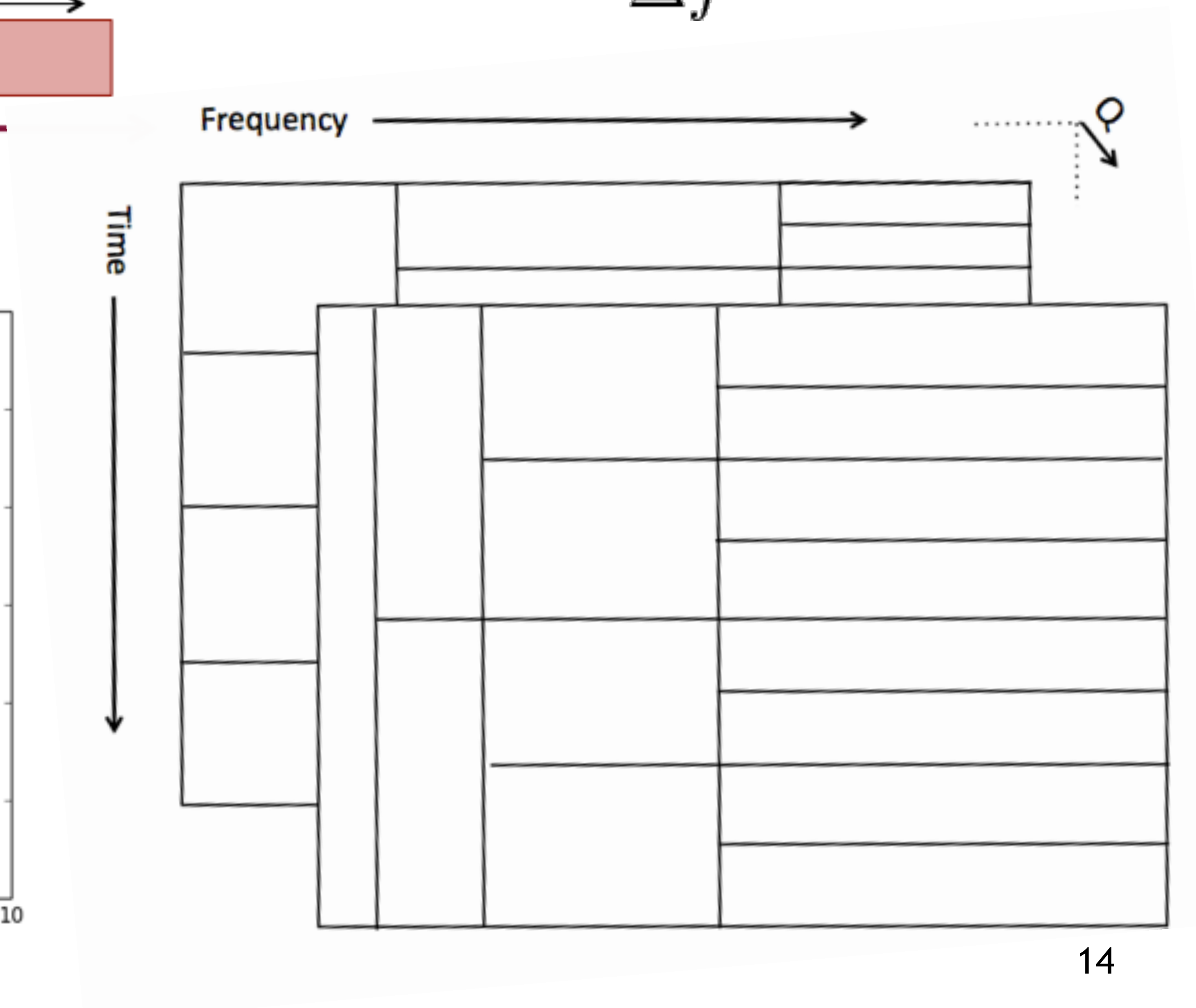
The Q transform

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

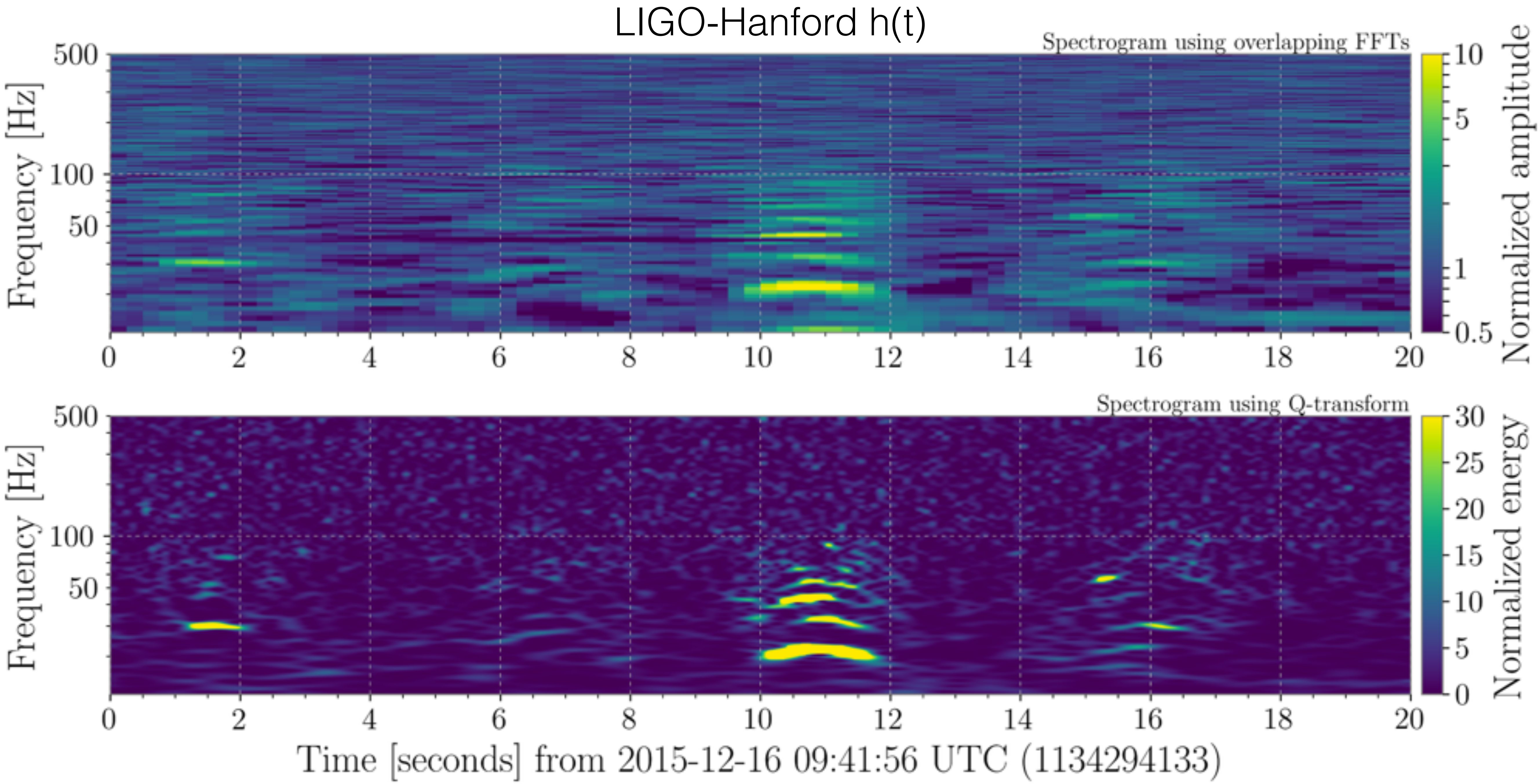


$$Q = \frac{f_0}{\Delta f}$$





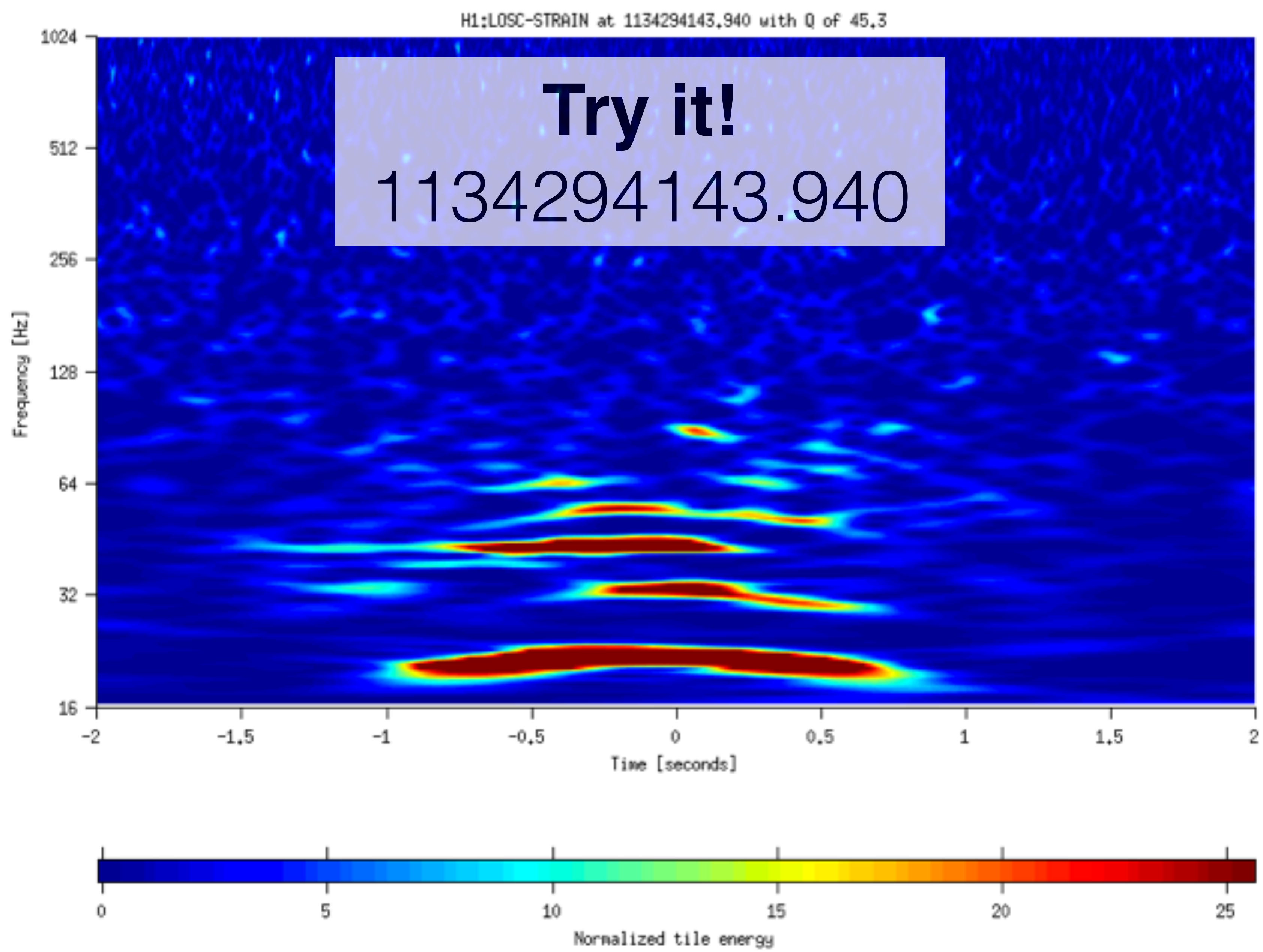
Time-frequency spectrograms



Made with GWpy by Duncan Macleod

Public resources for Q transforms

To generate a Q transform of open LIGO data, visit qscan.ligo.org



Coming soon! LIGO Data Viewer web for open LIGO and Virgo data

Will support: time series, spectra, spectrograms, Q transform, some filtering, band-limited root-mean-square

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

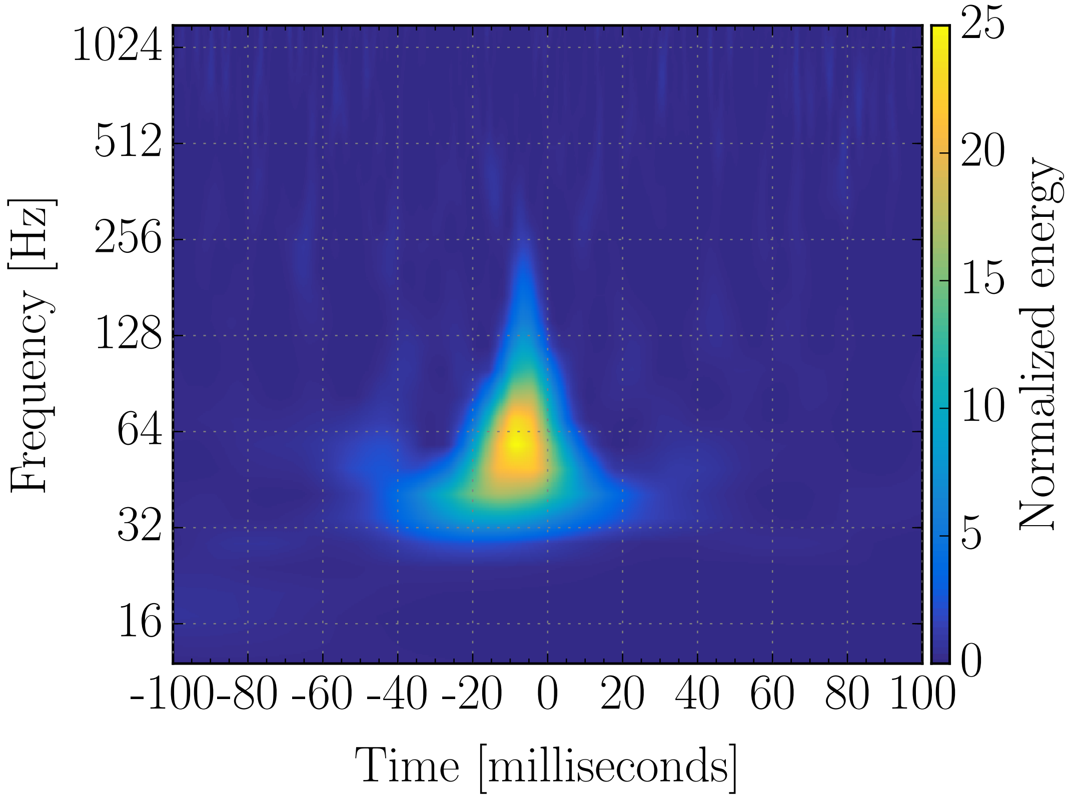
LIGO data is non-stationary!

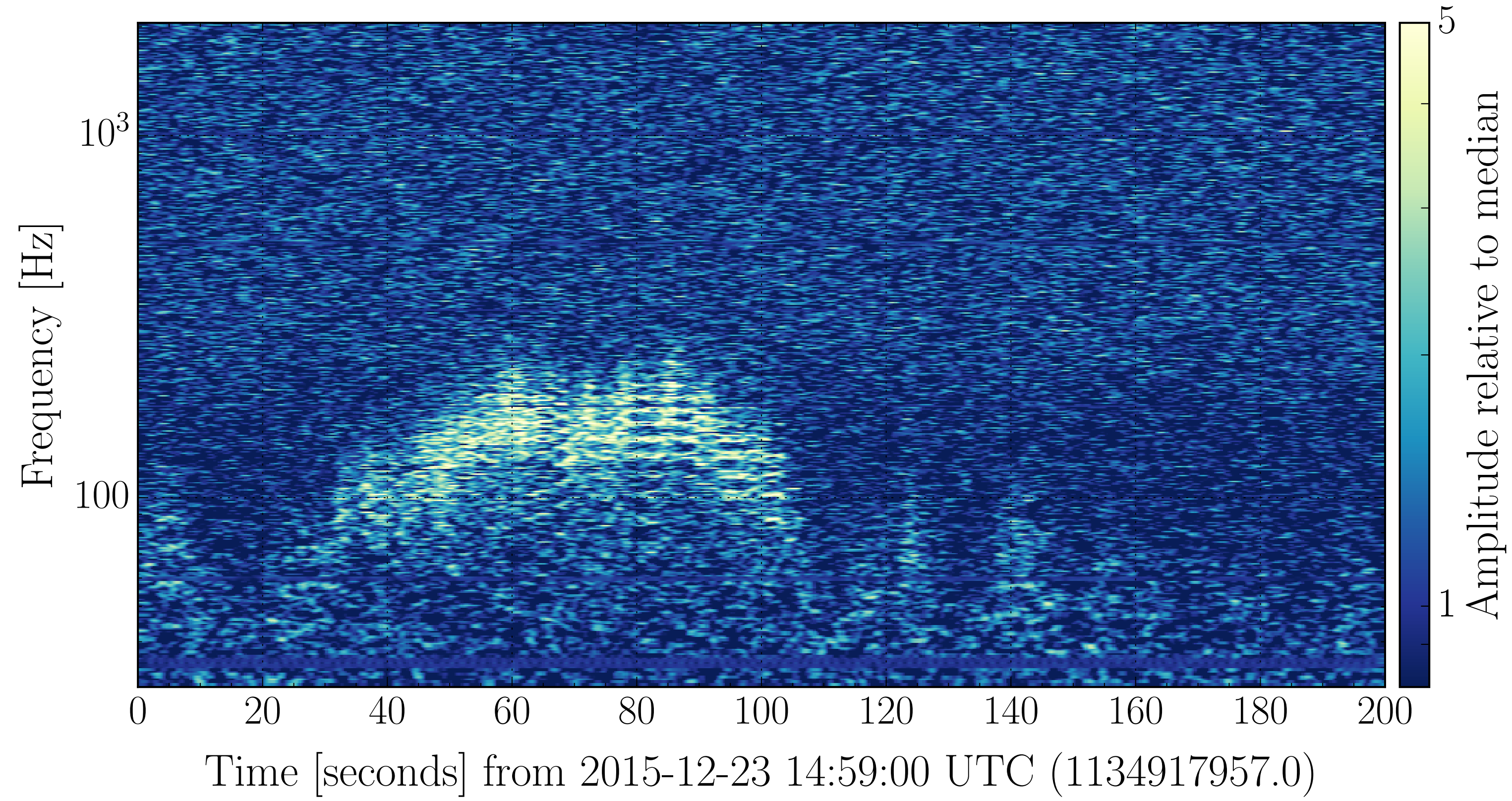
Blip glitches

- The biggest contributor to the transient GW search backgrounds
- Seen in both LIGO detectors (non-coincident) ~1x/hour
- No known correlation with instrument behavior or environment.

60-200 Hz non-stationary noise

- Pollutes LIGO-Livingston data in a critical frequency range (~50-500Hz)
- Longer duration (10s or 100s of seconds)
- Major contributor to CBC and burst backgrounds

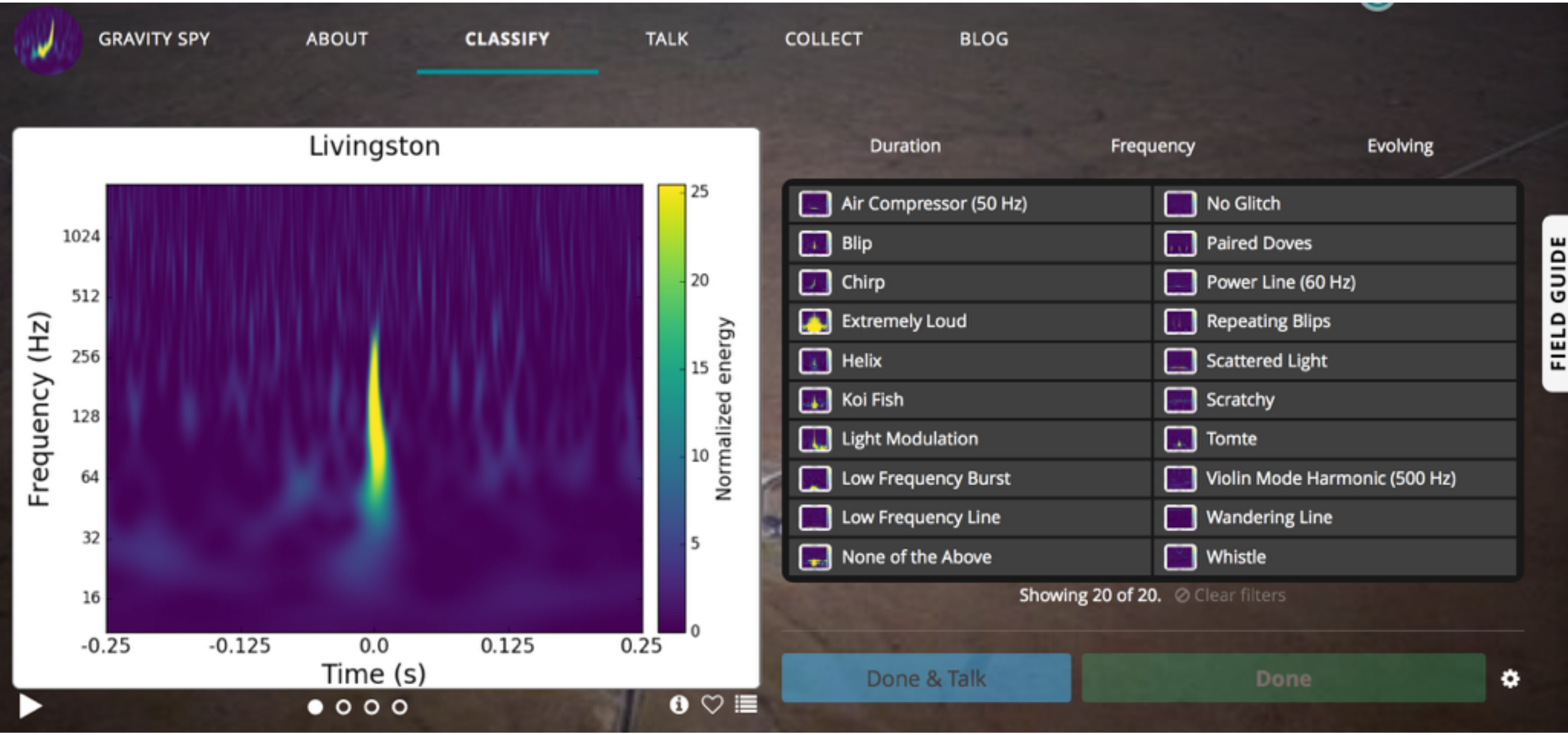




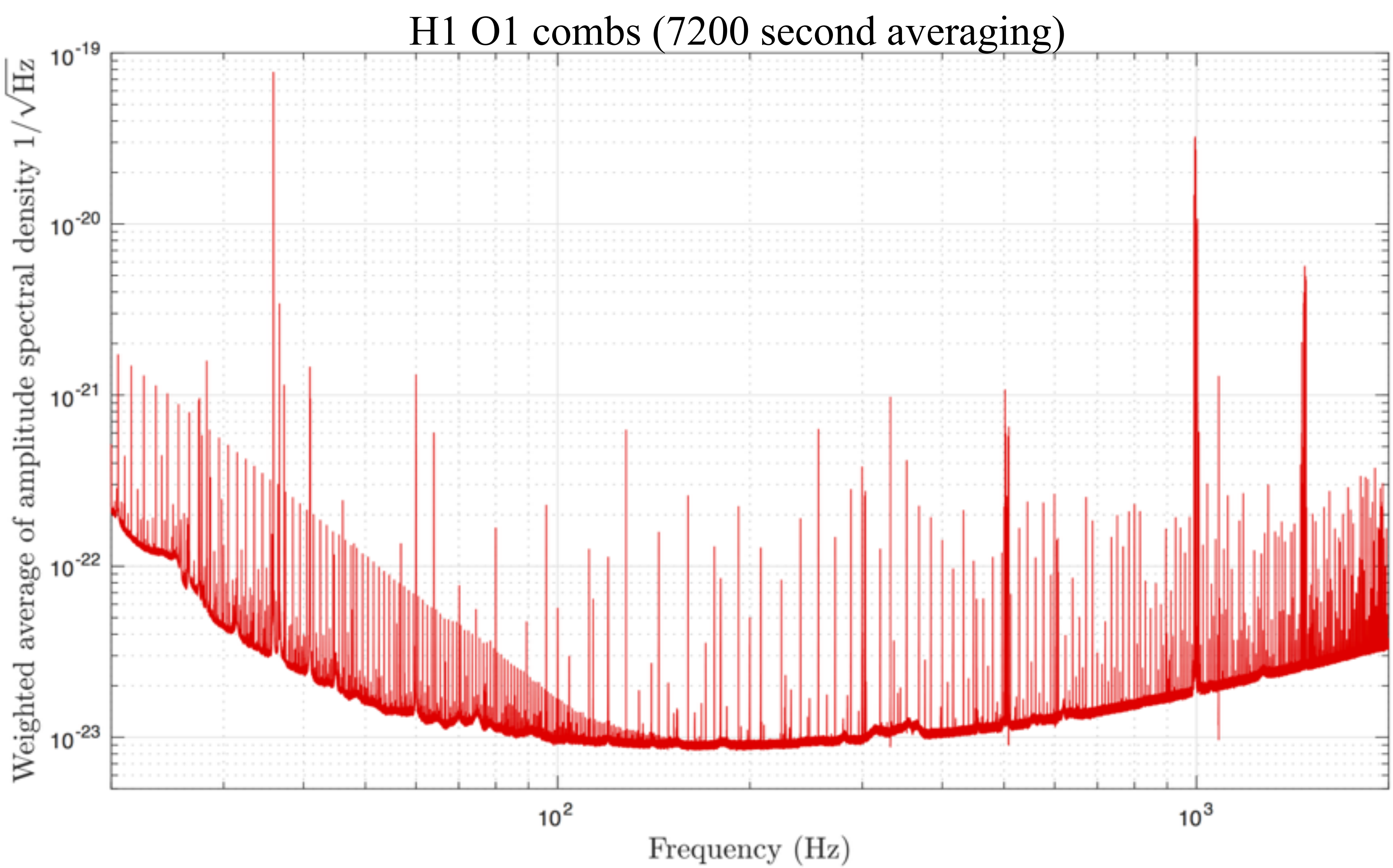
arXiv 1710.02185

A menagerie of common glitch types

gravityspy.org   Zevin et al, 2017, CQG



Combs of lines in LIGO data



O1 and O2 noise lines paper: Covas et al. (2017) arXiv 1801.07204
Instrumental lines catalog for LIGO-Hanford and LIGO Livingston: losc.ligo.org/o1speclines

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

O1 data quality information

Available via the LOSC

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 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 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.*

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.

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 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.

How are data quality segments defined?

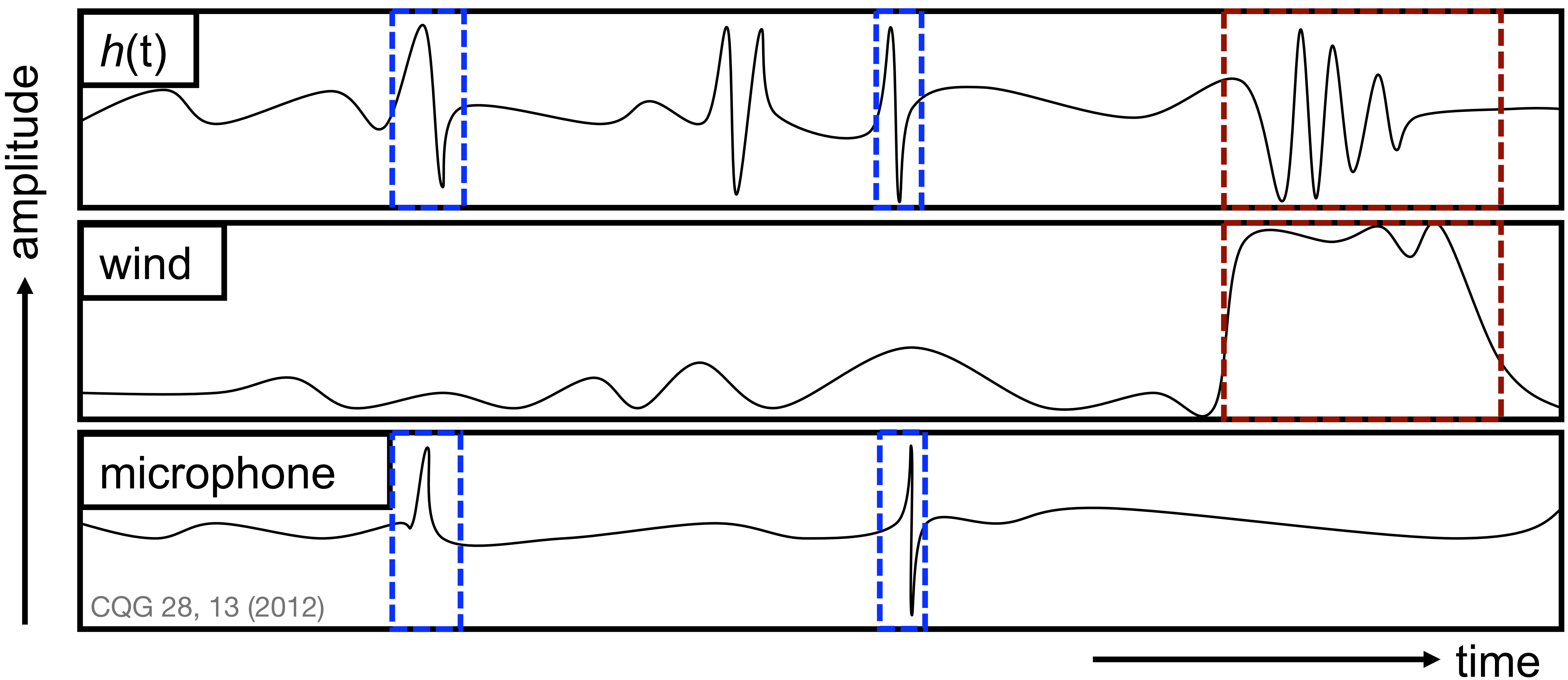
For O1 and prior analyses:

- Data quality vetoes required an *auxiliary witness*
- That auxiliary witness was required to be *safe*; to not be sensitive to changes in spacetime strain
- Veto segments were defined based on noise sources known to couple to h(t)
- Veto categories were determined for each type of search independently depending on noise contributors to that search's background
    - There are differences between CAT2 and CAT3 definitions between the burst and CBC searches

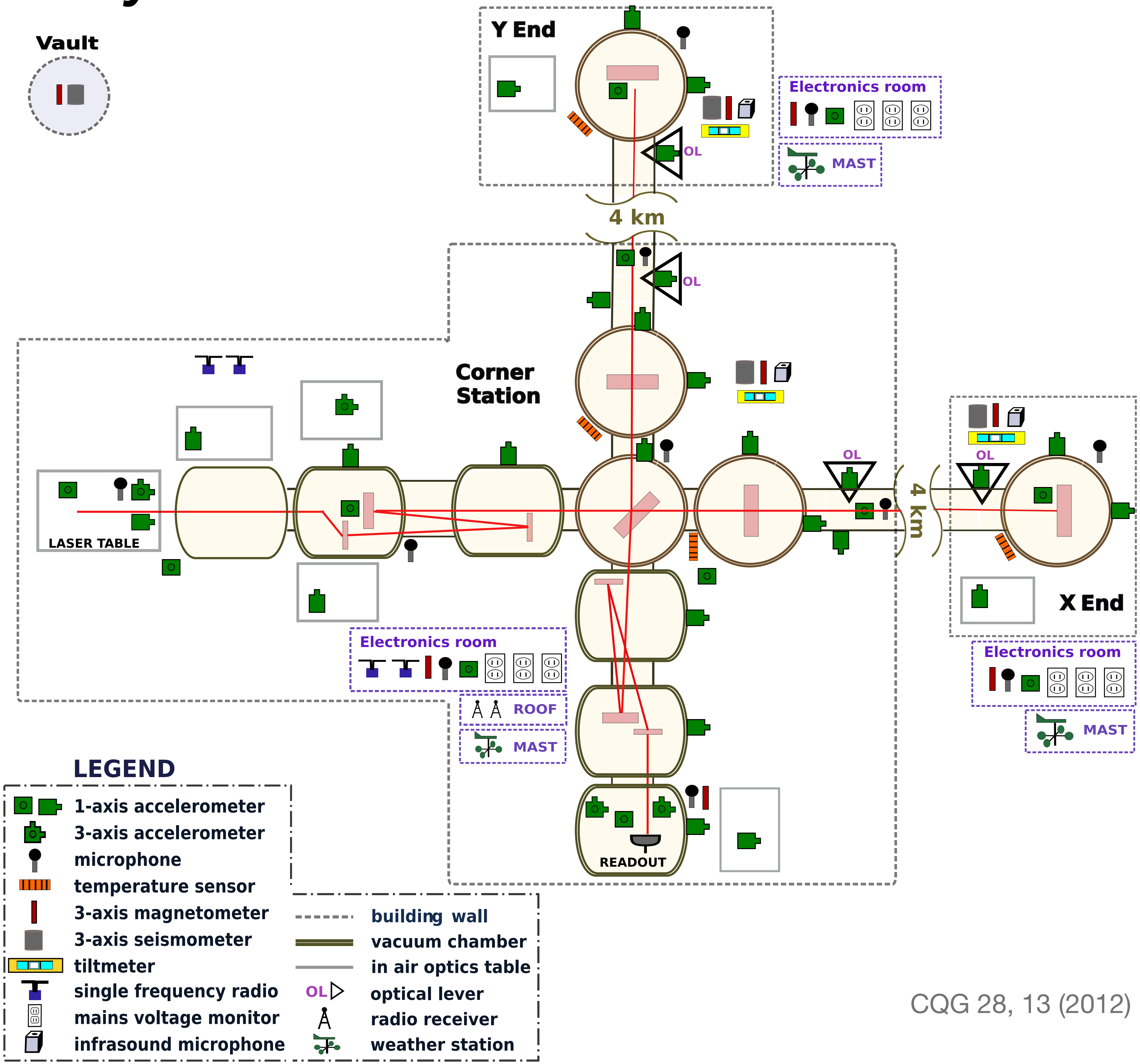
Auxiliary channels

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.

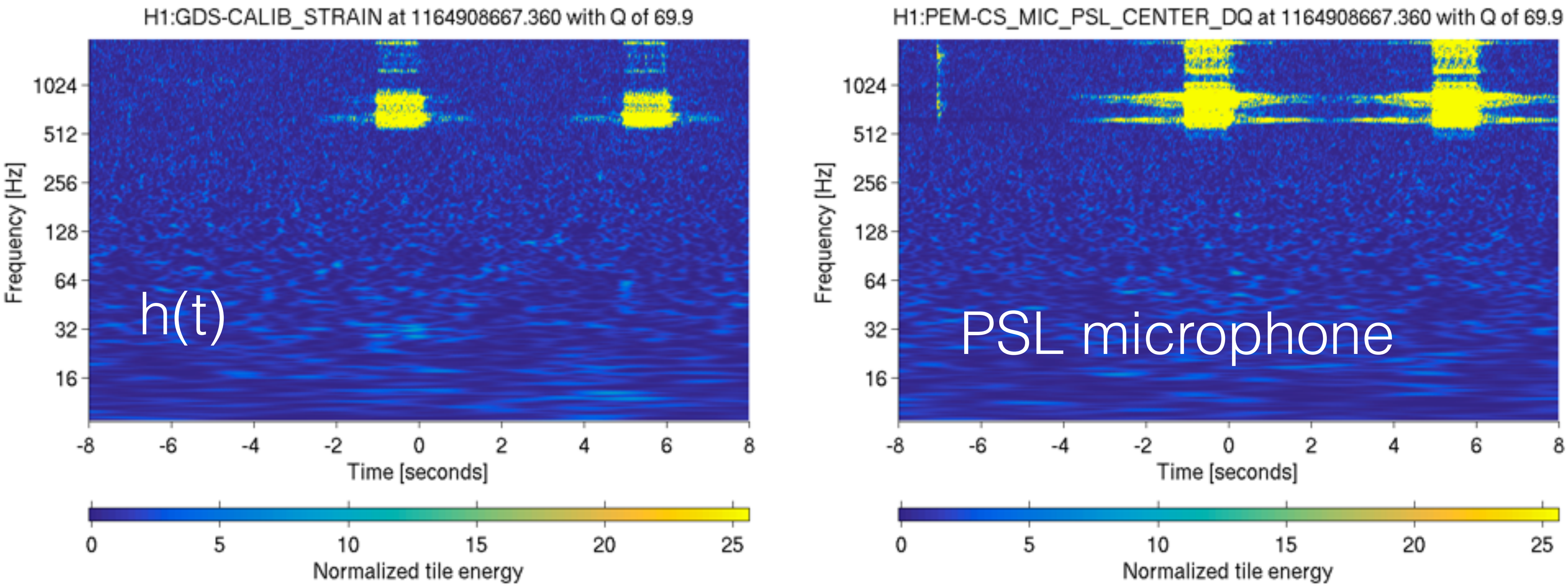


Physical environment channels

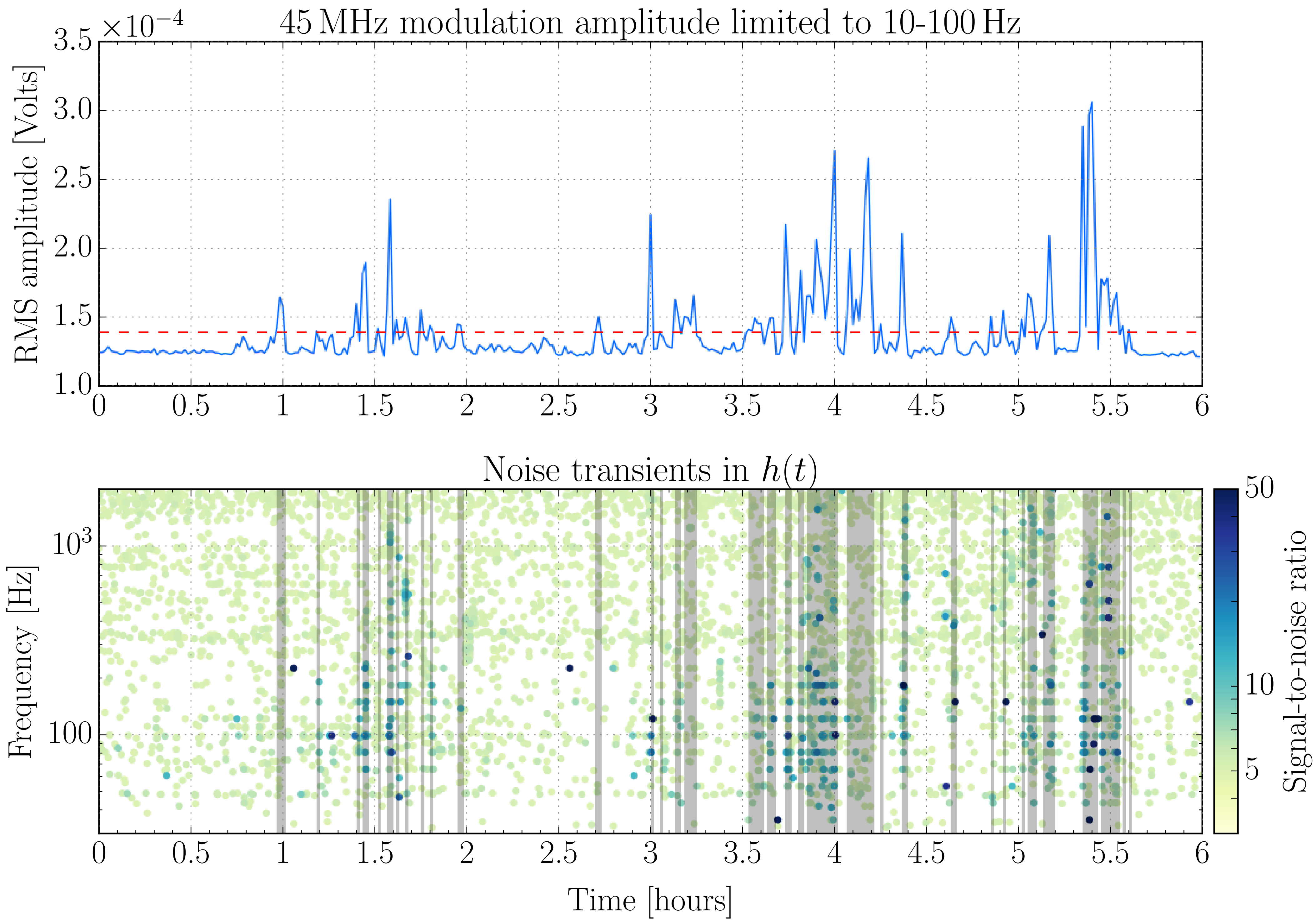


CQG 28, 13 (2012)

Laser glitches - h(t) vs. microphones



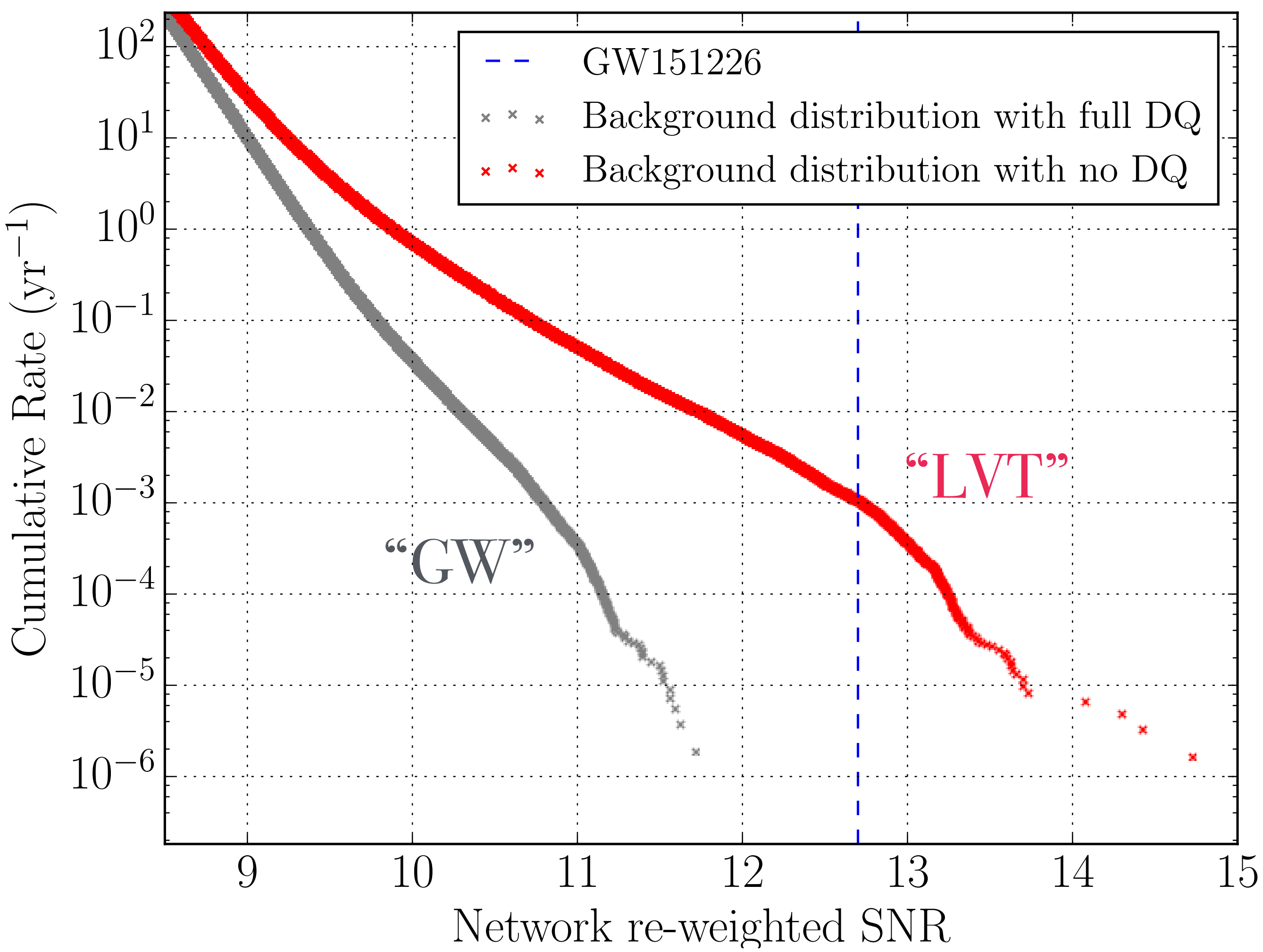
Example of a data quality veto in O1



LIGO-Virgo collaboration (2016) - arXiv 1602.03844

The impact of data quality vetoes

GW151226 analysis



LIGO-Virgo collaboration (2017) - arXiv 1710.02185

The false alarm rate of GW151226 **improves by a factor of 567**, from 1 in 320 years to 1 in 183000 years, **with interferometer data quality information!**

See Alex Nitz's talk for more on calculating event significance with PyCBC and analysis-specific mitigation methods.

Event validation

Is there noise or non-stationarity present that could account for the identified signal?

Is there noise or non-stationarity present that could interfere with accurate source property estimation?

Auxiliary sensors are considered a crucial test for event validation.

Note these channels are not currently released as open data.

Outline

What's in a LIGO data file?

What does LIGO strain data look like?

- Time domain
- Frequency domain
- Time-frequency representations

Data quality: noise artifacts in LIGO data

- Glitches
- Lines

Mitigating noise artifacts

- Data quality vetoes
- Analysis-dependent mitigation
- Event validation

Summary of resources and references

Useful data quality references

For glitches:
GW150914 Detchar 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
O1 lines catalog on the LOSC: losc.ligo.org/o1speclines/
GW150914 calibration paper: arXiv 1602.03845
Specific O2 event releases: calibration lines and 60 Hz lines removed

Resources:
Data quality information on the LOSC: losc.ligo.org/archive/dataset/O1/
Public Qscans: qscan.ligo.org
Coming soon! Public LIGO-DV web

Public interferometer status monitoring: https://losc.ligo.org/detector_status/

Spectral density estimation

The signal amplitude for each discrete frequency is equal to the square root of the signal "power"; thus units of 1/sqrt(Hz)

"Power" is proportional to the mean square of the signal for each discrete frequency (calculated using averaged FFTs)

$$\tilde{S}_{xx}(\omega) = \frac{(\Delta t)^2}{T} \left| \sum_{n=1}^{N} x_n e^{-i\omega n} \right|^2$$

For each discrete frequency w
For each time sample n (summed from 1 to N)
For signal value x at time sample n and frequency w