NRGW summer school, 2024.07.29~08.02

Data Quality (DQ) impact on GW Search

Kim, Young-Min (김영민) KASI(한국천문연구원)

Email: <u>ymkim@kasi.re.kr</u> <u>ymkim715@gmail.com</u>

Astrophysical GW Sources



Gravitational-Wave Detection Network



Lickaser Interferometer



What does LIGO data look like?



LIGO data in the frequency domain



Made with GWpy by Duncan Macleod. Code: <u>https://git.io/gwpy-ligo-scattering-animation</u> 0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

LIGO data in the frequency domain



Made with GWpy by Duncan Macleod. Code: <u>https://git.io/gwpy-ligo-scattering-animation</u> 0.5 second FFT; 5 averages covering 1.5 seconds; 50% overlap

PSD - averaging



Bartlett's Method (windowless)

https://www.researchgate.net/figure/Welchs-and-Bartletts-methods-for-power-spectraldensity-estimation-The-Bartletts_fig1_349283231 7

Detector Sensitivity



Features in GW data



Laura Nuttall in GW ODW #4, 2021

Q-transform



$$Y(t, t_0, f_0, Q) = \left(\frac{8\pi f_0^2}{Q^2}\right)^{\frac{1}{4}} \exp\left[\frac{-4\pi^2 f_0^2}{Q^2} \left(t - t_0\right)^2\right] \exp\left[-i2\pi f_0 \left(t - t_0\right)\right] \quad (3.11)$$

Q-transform

< ∆t >

or equivalently in the frequency domain:

$$\widetilde{Y}(f, t_0, f_0, Q) = \left(\frac{Q^2}{2\pi f_0^2}\right)^{\frac{1}{4}} \exp\left[\frac{-Q^2}{4f_0^2}(f - f_0)^2\right] \exp\left[-i2\pi t_0(f - f_0)\right]. \quad (3.12)$$



GWI708I7: Spectrograms in gwpy



GWI708I7: Q-transform in gwpy

1 ldata = TimeSeries.fetch_open_data('L1', *segment, verbose=True) $1 \lg = 1 \operatorname{data.q} \operatorname{transform}(\operatorname{frange}=(30, 500), \operatorname{qrange}=(100, 110))$ 2 plot = lq.plot() 3 ax = plot.gca() 4 ax.set_epoch(gps) 5 ax.set_yscale('log') 6 ax.colorbar(label="Normalised energy") $\times 10^4$ 500 - 1.4 - 1.2 Normalised energy Frequency [Hz] 100 0.4 50 0.2 0.0 -21 -18 -15 -12 -3 -30 -27 -24 -9 -6 0 Time [seconds] from 2017-08-17 12:41:04.4 UTC (1187008882.4)

GWI708I7: Q-transform in gwpy

1 plot.colorbars[0].mappable.set_clim(0,20)
2 plot.refresh()
3 plot



Detector Status in GWOSC



BNS Range

$$\varrho = \sqrt{4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)} df}$$
$$= \left(\frac{1 \text{ Mpc}}{D_{\text{eff}}}\right) \sqrt{4 \mathcal{A}_1^2 \text{ Mpc}} (M, \mu) \int_0^\infty \frac{f^{-7/3}}{S_n(f)} df \qquad (D1)$$

Horizon distance

Expected SNR

$$D_{\rm hor} = \frac{1 \,\,{\rm Mpc}}{\varrho} \sqrt{4 \,\mathcal{A}_{1\,\,{\rm Mpc}}^2(M,\,\mu) \int_0^\infty \frac{f^{-7/3}}{S_n(f)} df}, \qquad ({\rm D2})$$

snr=8, m1=m2=1.4 Msun, mu=0.7 Msun

BNS range (Sense-monitor range)

PHYSICAL REVIEW D 85, 122006 (2012)

Example - BNS range by GWPy

First, we need to load some data. We can **fetch** the public data around the GW170817 BNS merger:

from gwpy.timeseries import TimeSeries
h1 = TimeSeries.fetch_open_data('H1', 1187006834, 1187010930, tag='C02')
l1 = TimeSeries.fetch_open_data('L1', 1187006834, 1187010930, tag='C02')

Then, we can measure the inspiral range directly:

```
from gwpy.astro import range_timeseries
h1range = range_timeseries(h1, 30, fftlength=4, fmin=10)
l1range = range_timeseries(l1, 30, fftlength=4, fmin=10)
```

We can now plot these trends to see the variation in LIGO sensitivity over an hour or so surrounding GW170817:

```
plot = h1range.plot(
    label='LIGO-Hanford', color='gwpy:ligo-hanford', figsize=(12, 5))
ax = plot.gca()
ax.plot(l1range, label='LIGO-Livingston', color='gwpy:ligo-livingston')
ax.set_ylabel('Angle-averaged sensitive distance [Mpc]')
ax.set_title('LIGO sensitivity to BNS around GW170817')
ax.set_epoch(1187008882) # <- set 0 on plot to GW170817
ax.legend()
plot.show() https://gwpy.github.io/docs/latest/examples/miscellaneous/range-timeseries.html</pre>
```

Example - BNS range by GWPy

First, we need to load some data. We can **fetch** the public data around the GW170817 BNS merger:





plot.show()

https://gwpy.github.io/docs/latest/examples/miscellaneous/range-timeseries.html

Noise backgrounds

Complicated noise curves

Many lines in the data, not such an issue for transient searches, but can be an issue for continuous wave searches

To an okay approximation, the detector data is colored Gaussian noise – standard Gaussian noise just with certain frequencies louder than others



Image from Abbott et al (2020) GWTC-2 2010.14527

Calibrated Strain noise spectral lines



Source: https://losc.ligo.org/events/GW150914/

Spectral Lines - official info.

O1 Instrumental Lines

The plot below shows the amplitude spectral density (ASD) of the strain noise in the H1 and L1 Advanced LIGO detectors, during a "typical" time in the O1 run. The plot shows frequency [Hz] on the X-axis, and the ASD value [1 / sqrt(Hz)] on the y-axis. The first thing to note is that the data are not calibrated or valid below 10 Hz or above 5 kHz (and the data sampled at 4096 Hz are not valid above 2000 Hz).

The spectra reveal a large number of "lines" due to instrumental artifacts:



O1: <u>https://www.gw-openscience.org/o1speclines/</u>

O2: <u>https://www.gw-openscience.org/o2speclines/</u>

O3a : https://www.gw-openscience.org/O3/o3aspeclines/

Noise backgrounds

Non-stationarity

The detector sensitivity is not constant, this can happen rapidly or slowly



LIGO Sources



Credit: Bohn, Hébert, Throwe, SXS

Coalescing Binary Systems

• Black hole – black hole

•Black hole – neutron star

Neutron star – neutron star

(modeled waveform)



Credit: Chandra X-ray Observatory

Transient 'Burst' Sources

- asymmetric core collapse supernovae
- cosmic strings
- ???

(Unmodeled waveform)



Credit: Planck Collaboration

Stochastic Background

- residue of the Big Bang
- incoherent sum of unresolved 'point' sources

(stochastic, incoherent noise background)



Continuous Sources

• Spinning neutron stars

(monotone waveform)

In D. Reitze's presentation in LIGO ODW #1, 2018

The Astrophysical Sources of GWs (I)

Transient sources



Credit: Albert Einstein Institute (AEI)



Credit: Chandra X-ray Observatory

Compact Binary Coalescence (modeled waveform)

Burst sources (un-modeled waveform)

The Astrophysical Sources of GWs (1)

Transient sources







Compact Binary Coalescence (modeled waveform) Burst sources (un-modeled waveform)

Non-Gaussian Transient Noises



Bahaadini et al. (2018)

The Astrophysical Sources of GWs (2)

non-Transient sources



Credit: Plank Collaboration

Stochastic Background



Credit: Casey Reed, Penn State

Continuous Sources

The Astrophysical Sources of GWs (2)

non-Transient sources



Credit: Casey Reed, Penn State

Stochastic Background

Continuous Sources

Spectral Lines and Combs



OI and O2 noise lines paper: Covas et al. (2017) arxiv: 1801.07204 Instrumental Lines catalog: www.gw-openscience.org/o1speclines

Detection limitation by Noises

LIGO

Non-Gaussian Transient Nosies : Glitches

Spectral lines : electrical or mechanical resonances



Credit: Bohn, Hébert, Throwe, SXS

Coalescing Binary Systems

• Black hole – black hole

•Black hole – neutron star

 Neutron star – neutron star (modeled waveform) • asyn collap • cosn • ???

Transient 'Burst' Sources

• asymmetric core collapse supernovae

cosmic strings

(Unmodeled waveform)

Credit: Chandra X-ray Observatory

Stor Bac • res • ince unres source

Credit: Planck Collaboration

Stochastic Background

- residue of the Big Bang
- incoherent sum of unresolved 'point' sources

(stochastic, incoherent noise background)

Continuous Sources

Spinning neutron stars

(monotone waveform)

Credit: Casey Reed, Penn State

Posteriors depending on DET status



Auxiliary Channels for DQ



Physical Environment Channels



Correlations with 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.



Laura Nuttall in GW ODW #4, 2021

Veto



Data Quality Flags: exclude periods of data for known noises **Data Quality Triggers**: short duration vetoes generated by algorithms that identify significant statistical correlation between a transient in h(t) and transient noise in auxiliary channels

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.

Laura Nuttall in GW ODW #4, 2021

Example of a data quality veto in O2





Gravitational Wave Open Science Center

A	Data - Software - Online Tools - About GWOSC -			
	Strain Data	Auxiliary Channel Three Hour Release		
Event Portal Timelines		Data Set A large number of sensors are used to record the state of the LIGO instruments and their enviornment. This data release contains sensor data recorded in around 500 channels at each LIGO site. These data represent three hours of time centered on GW170814 (GPS		
		1186736512 — 1186747264). Strain data from the same period are available in the O2 Data Release.		
	Auxiliary Channels Low Latency Alerts	Download Data The data are available as down-sampled HDF5 files [19 GB], or full sample rate GWF files [68 GB]:		
		♦ HDF5 Data ♦ GWF Data		
		Data may also be accessed from a network data server (NDS2) using the NDS2 client or GWpy:		
		<pre>from gwpy.timeseries import TimeSeries data = TimeSeries.fetch('L1:LSC-DARM_OUT_DQ', start=1186741850, end=1186741870, host='losc-nds.ligo.org')</pre>		
		See the NDS2 Example Code for details.		
		Example Software Example software is available in an associated git repo. To work with GWF files, see the software page.		
Channel Descriptions This data set is designed to be used for subtracting noise sources - especially contr are those most likley to include a coupling to the gravitational wave strain channel,		Channel Descriptions This data set is designed to be used for subtracting noise sources - especially controls noise - from LIGO data. Channels included in this set are those most likley to include a coupling to the gravitational wave strain channel, and so are possible sources of noise.		
		The Channel List shows each channel with a few properties:		
		 Channel name Desired sample rate: The sampling rate in the down-sampled, HDF5 data 		
		 Notes: A brief note explaining the meaning of the data in the channel Calibration: Where available, a calibration factor is included. Most channels are not calibrated. 		

• Units: Where available, the units corresponding to the calibration factor

In some cases, data for a given channel may not be available. These are marked "invalid" in the HDF5 files, and the corresponding channel may be absent from the GWF files. Unavailable channels may correspond to sensors that are not present at a particular site or not operational at a particular time.

Veto



Data Quality Flags: exclude periods of data for known noises **Data Quality Triggers**: short duration vetoes generated by algorithms that identify significant statistical correlation between a transient in h(t) and transient noise in auxiliary channels

Counting Experiment





Poisson statistics

- Poisson distribution expresses probability of a number of independent events occurring in a given time period
- apply to coincidence
- Definitions
 - $N_{de} =$ **number** of triggers in DARM ERR channel
 - N_n = number of triggers in auxilliary channel n
 - T_{win} = full time window centered on auxilliary channel trigger
 - T_{tot} = total live-time analyzed
- From these calculate mean
 number of expected coincidences

Probability Density Function

$$PDF_{poi}(\mu, x') = \frac{\mu^{x'} e^{-\mu}}{x'!},$$

Mean number of coincidence

$$\mu = \frac{N_{de} N_n T_{win}}{T_{tot}}$$

Statistical Significance



Statistical Significance



Veto Algorithms (I)

I. Use-Percentage Veto (UPV)



2. Hierarchical Veto (Hveto)

$$S = -\log_{10} \sum_{k=n}^{\infty} \left[\frac{\mu^k e^{-\mu}}{k!} \right]$$
$$\mu = \frac{N_{main_tot} N_{aux_tot} T_{win}}{T_{tot}}$$

Journal of Physics: Conference Series 243, 012005 (2010); 14th GWDAW

n : the number of coincidences T_win : full width of coincidence time window T_tot : a given total analysis time

Class. Quantum. Gav. 28, 235005 (2011)

Veto Algorithms (2)



Class. Quantum. Gav. 30, 155010 (2013)

hVeto



n : the number of coincidences T_win : full width of coincidence time window T_tot : a given total analysis time



Non-Gaussian Transient Noises



Bahaadini et al. (2018)

Glitch Classification by ML



Vectorized information

MLA application to DQ

- I. Ordered Veto List (OVL) + 3 Machine Learning Algorithms
 - application to hundreds of channels among 200,000 auxiliary channels



Gravity Spy

Non-Gaussian Glitches



Gravity Spy

downsample: 140*170 —> merged view images (0.5, 1.0, 2.0, 4.0s)



+ Support Vector Machine (SVM) Ensemble Learning samples: 8583 train: 6008 validation: 1288 Test: 1287

Bahaadini et al. (2018)

CAGMon

- I. Developers : J. J. Oh (오정근), P. J. Jung (정필종)
- 2. Wiki: https://kgwg.nims.re.kr/wiki/DetChar/CAGMon
- 3. Code : <u>https://github.com/pjjung/cagmon</u>



T.Washimi et al 2021 JINST 16 P07033



0.08

0.06

j ₩ 0.04

0.06

o.05

5 0.04

₿ 0.03-

0.02

0.06

0.05

0.04

0.03

Active

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.

DQ Category

Bit	Short Name	Description		
Data Quality Bits				
0	DATA	data present		
1	CBC_CATI	passes the cbc CAT1 test		
2	CBC_CAT2	passes cbc CAT2 test		
3	CBC_CAT3	passes cbc CAT3 test		
4	BURST_CATI	passes burst CATI 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		

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

DQ Category

Bit Short Name		Description
Data Qua	ality Bits	
0	DATA	https://gwosc.org/tutorial03/
1	CBC_CAT1	⁸ Good Data
2	CBC_CAT2	7- BURST_CAT1
3	CBC_CAT3	6 DATA
4	BURST_CAT1	5
5	BURST_CAT2	
6	BURST_CAT3	4
Injection	Bits	3
0	NO_CBC_HW_INJ	2
1	NO_BURST_HW_INJ	1
2	NO_DETCHAR_HW_INJ	
3	NO_CW_HW_INJ	
4	NO_STOCH_HW_INJ	-1 0 500 1000 1500 2000 2500 3000 3500 4000 Time (s)

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



Gravitational Wave Open Science Center



Segment Information

- I. segment DB : <u>https://segments.ligo.org</u>
 - query an available segment to segment DB
- 2. Public segment information in GWOSC (<u>www.gwosc.org</u>)

Timeline The vertical axis indicates the fraction of time a flag is on during each "Sample time".





Gravitational Wave Open Science Center





Gravitational Wave Open Science Center



Data Quality Impact on GW searches



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!**

LIGO-Virgo collaboration (2017) - arXiv 1710.02185

Thank you for your attention.

ATT