DEEPCLEAN: ML-ASSISTED NOISE REGRESSION IN GRAVITATIONAL WAVE DETECTORS

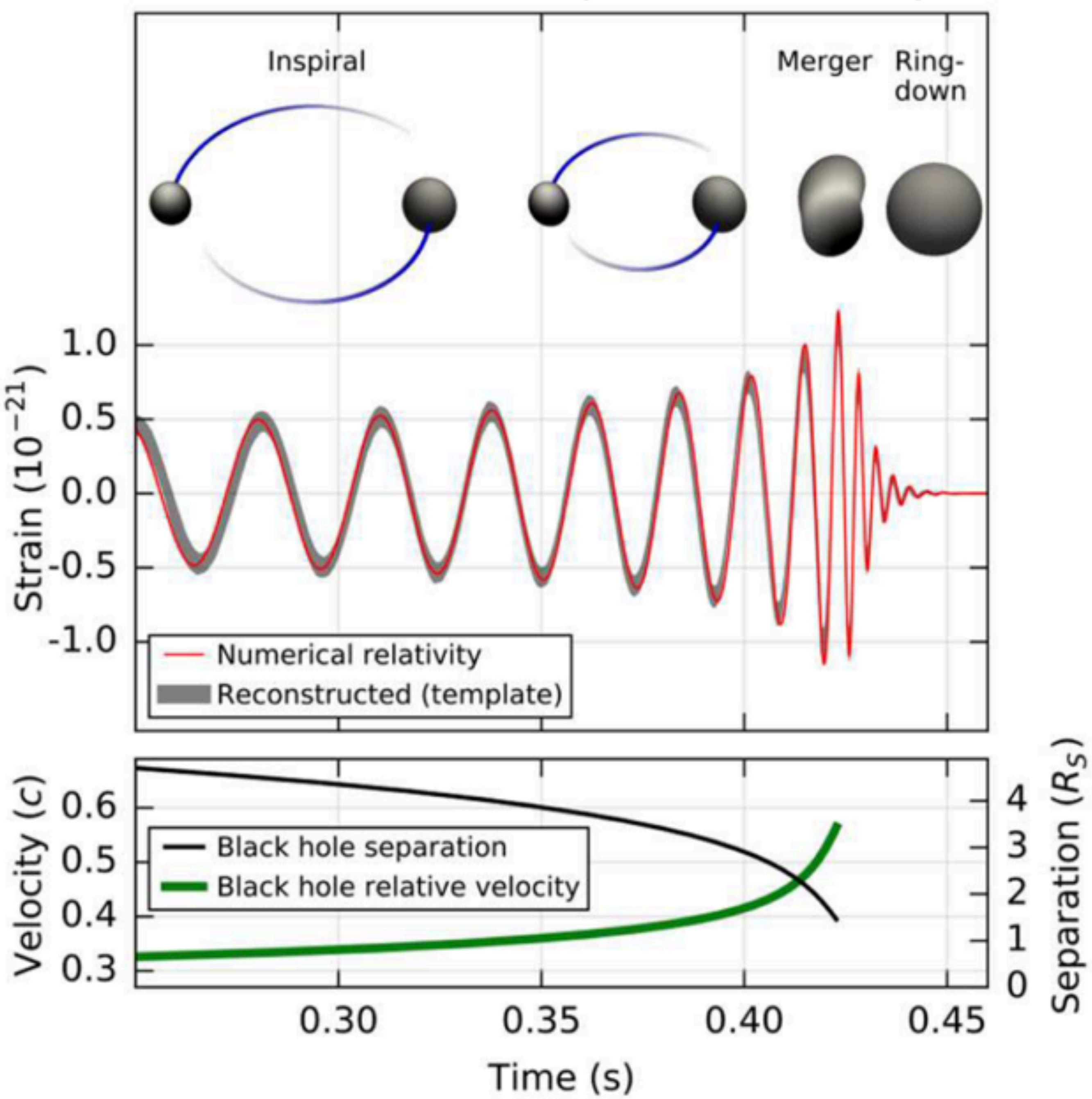
Muhammed Saleem,
University of Minnesota

A3D3 Fast ML Meeting, 3-6 Oct 2022, SMU, Dallas

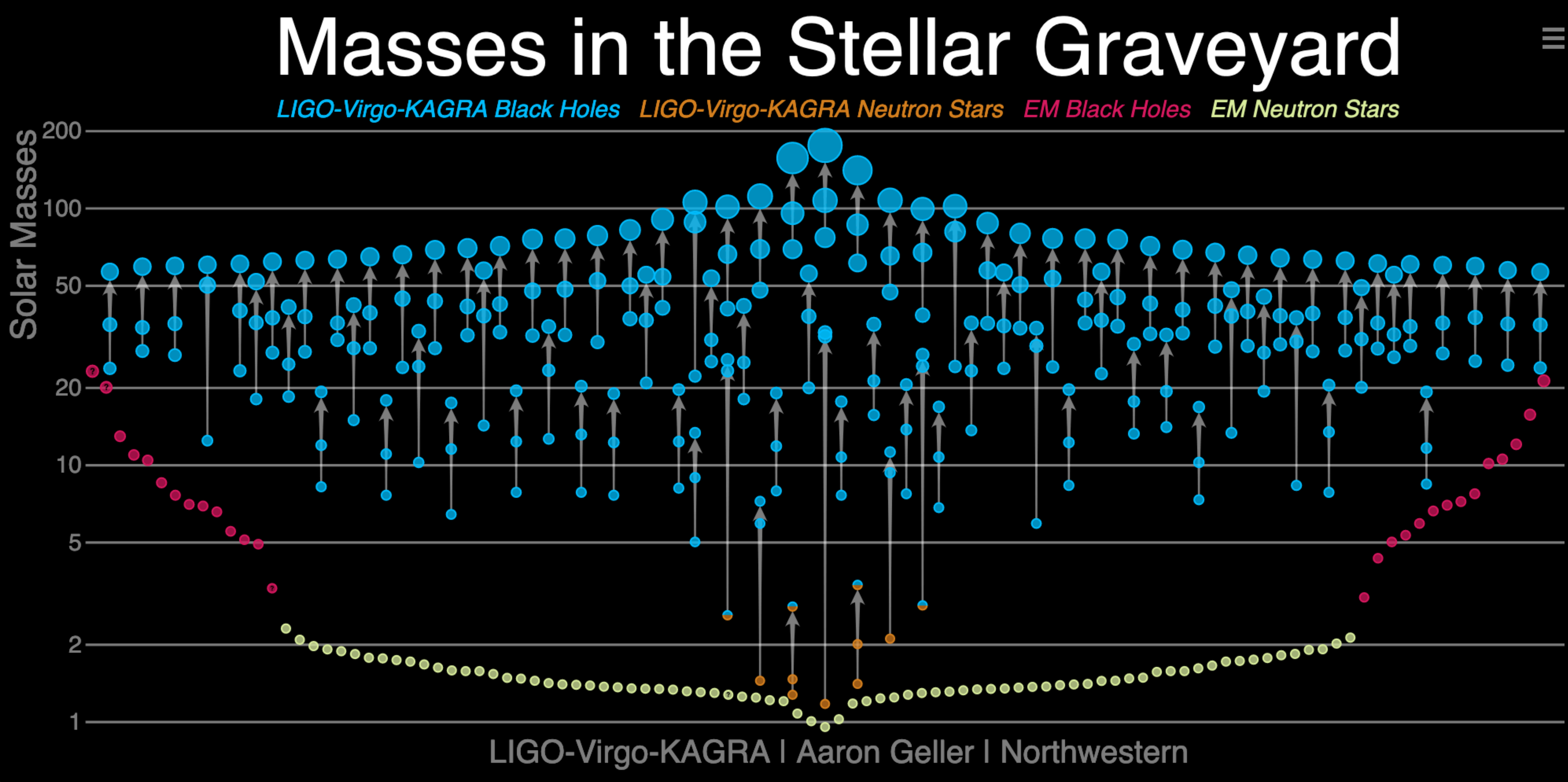
Collaborators: Alec Gunny, Li-Cheng Yang, Chia-Jui Chou, Michael Coughlin, Dylan Rankin, Erik Katsavounidis, Phil Haris et al.

GRAVITATIONAL WAVES (GW)

Figure: LIGO/Virgo

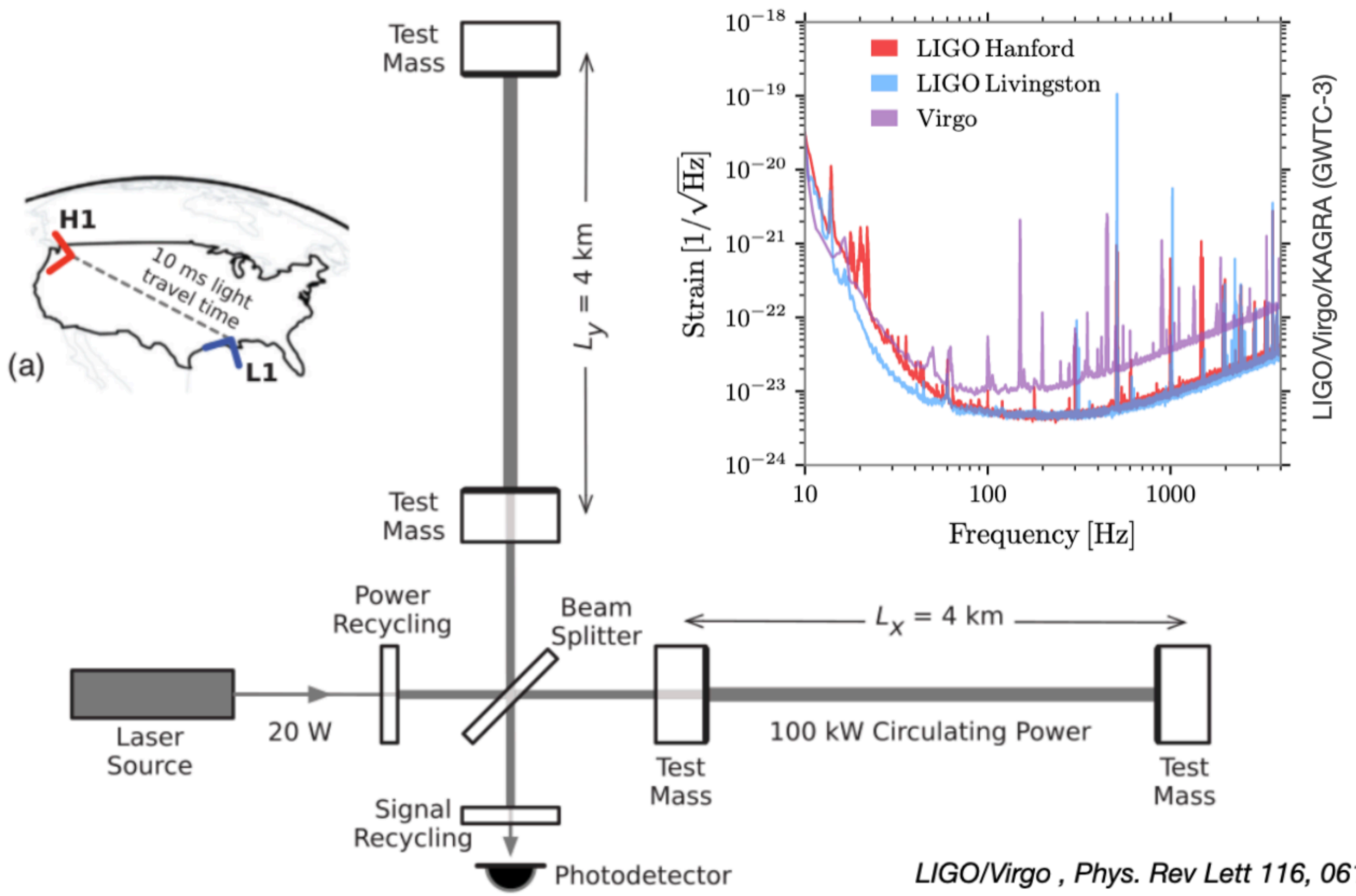


- Ripples in spacetime caused by massive objects moving with extreme acceleration (eg: pair of neutron stars and/or blackholes orbiting each other)
- Predicted by Albert Einstein in 1916 in his General theory of Relativity
- First direct detection made in 2015 by the two LIGO detectors at Hanford and Livingston, from a pair of colliding blackholes
- When they pass through earth, they produce strain $\sim 10^{-21}$ (producing displacements that are several thousands of times smaller than an atomic nucleus)
- Needs extremely sensitive precision instruments for detecting them.



So far, about one hundred merging blackholes/neutron stars observed in the GW window

Gravitational wave detectors



LIGO/Virgo , Phys. Rev Lett 116, 061102 (2016)

- A Michelson interferometer at heart, but several key differences
- Kilometre-scale, perpendicular arms with powerful laser beams producing interference patterns, due to distortions produced by the passing GWs

Challenges in detection: Signals buried in the noise

- The output reconstructed from an interferometer contains

$$h(t) = s(t) + n(t)$$

Possible GW signal ↗ $s(t)$ Detector noise ↖ $n(t)$

- Objective: To recover $s(t)$ with best possible signal-to-noise ratio by minimising the noise $n(t)$

- Many signals are below the noise $\Longrightarrow$ noise reduction can lead to many more detections

REMOVABLE AND NON-REMOVABLE NOISES

$$n(t) = n_{nw}(t) + n_{w}(t)$$

- Non-removable (fundamental noise)
- Budgeted by system design
- Eg: photon shot noise, thermal noise
- Can be reduced only with upgraded design and technology

- Source of noise witnessed by dedicated system monitors (witness sensors)
- Environmental contamination or technical noise eg: noise arising from the control of suspended optics

WITNESS CHANNELS

- At the interferometer sites, apart from the strain channel, there are thousands of complementary channels that record the random processes of environmental and instrumental origins

- known as witness sensors - the timeseries denoted by $\{w_i(t)\}$

- The noise $n_w(t)$ is the collective contribution from $\{w(t)\}$

- Schematically,

$$n_w(t) = \mathscr{T}\left(\{w_i(t)\}\right)$$

where $\mathscr{T}$ is some activation function representing non-linear or non-stationary coupling of the output of witness channels to the strain channel.

DEEPCLEAN: A NEURAL NETWORK TO PREDICT $n_w(t)$

$h(t)$

$\{w_i(t)\}$

$$n_w(t) = \mathscr{F}(\{w_i\}; \vec{\theta})$$

$$r(t) = h(t) - n_w(t)$$

$r(t)$

Residual strain after noise subtraction

- $\vec{\theta}$ → trained weights of the neural network

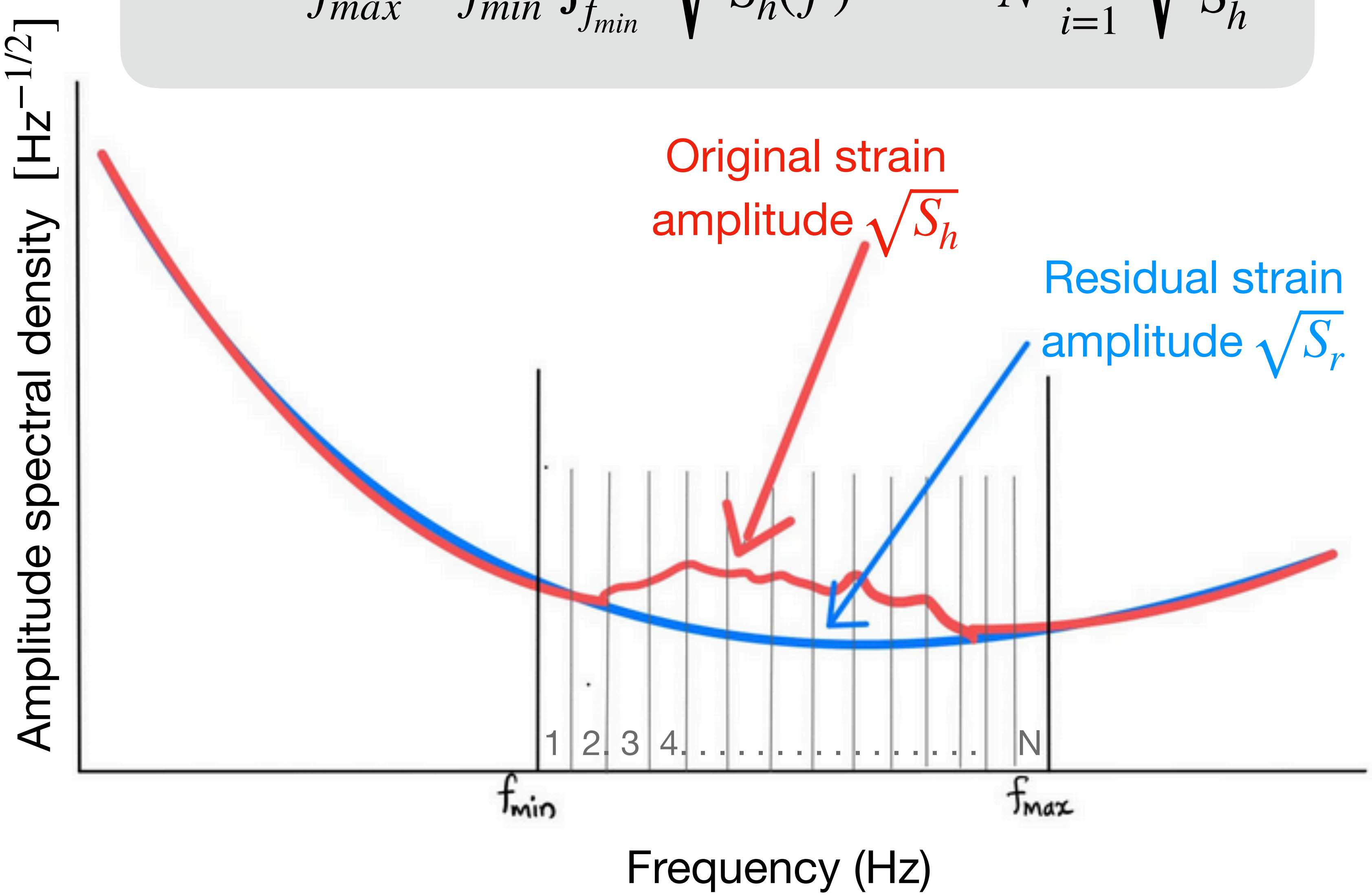
$$\vec{\theta} = \min_{\vec{\theta}'} \left[J \left(h(t), \mathscr{F}(\{w_i(t)\}; \vec{\theta}') \right) \right]$$

where J is some appropriate loss function

Ormiston et al (2020)

Loss function in terms of Amplitude Spectral Density (ASD)

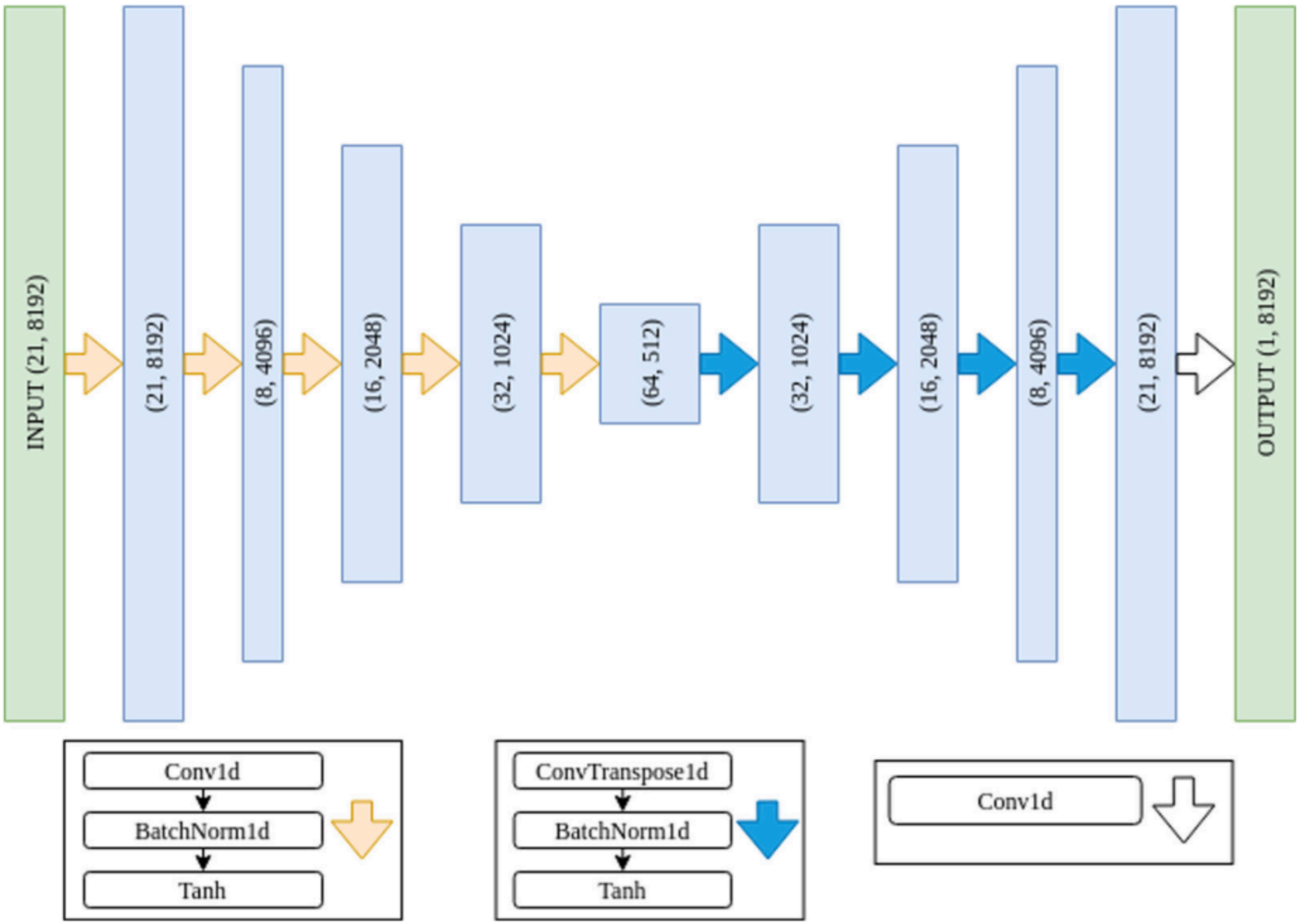
$$J_{asd} = \frac{1}{f_{max} - f_{min}} \int_{f_{min}}^{f_{max}} \sqrt{\frac{S_r(f)}{S_h(f)}}\, df = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{S_r^{(i)}}{S_h^{(i)}}}$$



- Minimises the ASD ratio averaged over all the bins in the desired frequency range.
- Ratio of residual ASD to the original ASD
- Involves FFT → the minimum length of the time-domain data dictated by the resolution of the features we are looking for.

Ormiston et al (2020)

THE DEEPCLEAN ARCHITECTURE



Fully convolutional auto-encoder mapping the witness channel data $\{w_i(t)\}$ into the noise prediction $n_w(t)$

Input: 8 s time-series from 21 witness channels sampled at 1024 Hz

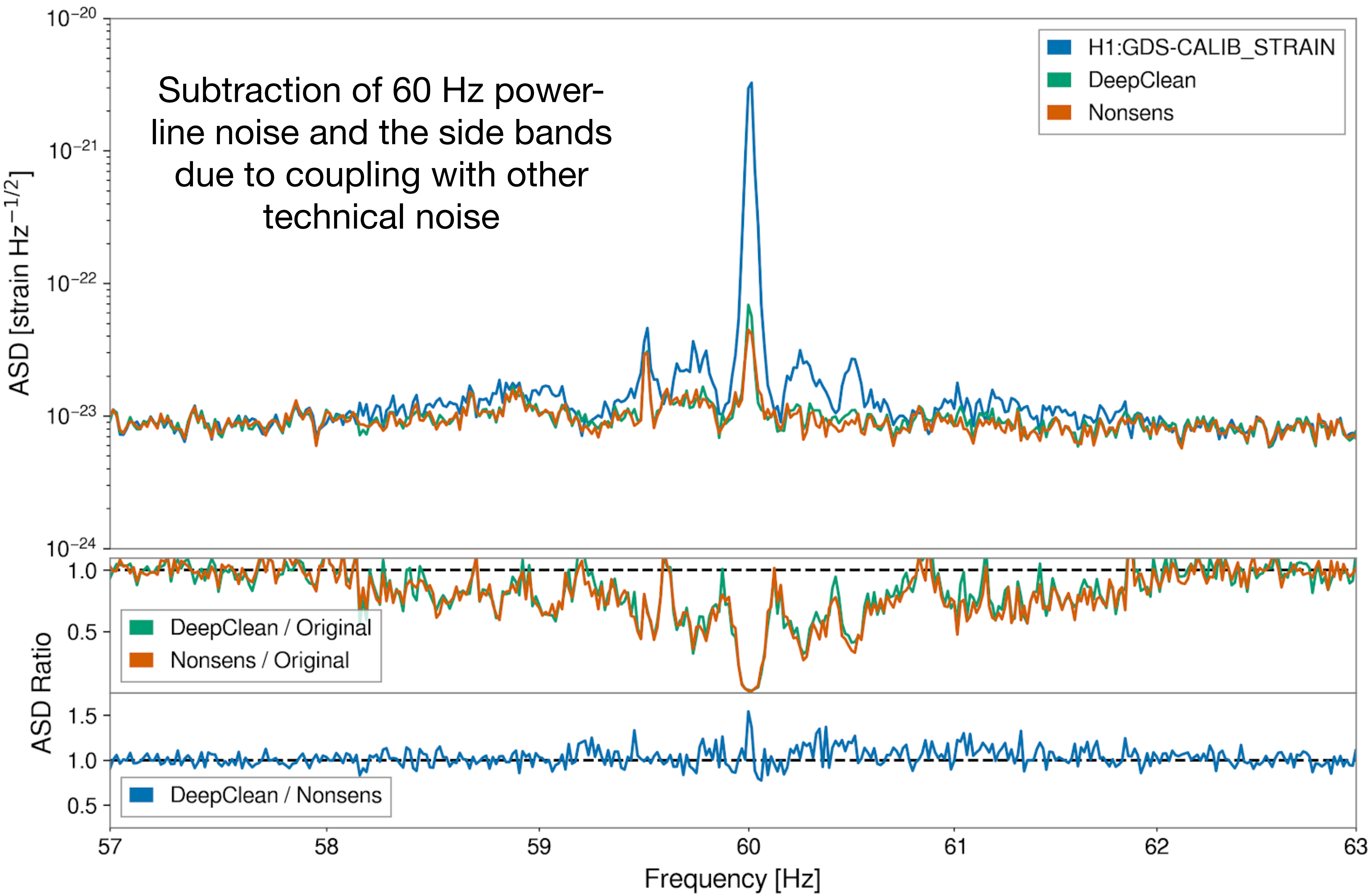
Output: Noise prediction for 8s sampled at 1024 Hz

Ormiston et al (2020)

TRAINING DEEPCLEAN

- Preprocess: normalize (zero mean and unit variance) and bandpass to the desired frequency band.
- Data sent to the network in mini-batches with one batch typically having 32 samples
- Loss function computed for the network prediction by averaging over the mini-batch.
- Loss is minimized using ADAM (the Gradient Descent algorithm) using both Forward pass and backward pass

Performance on LIGO data from third observing run (2019)



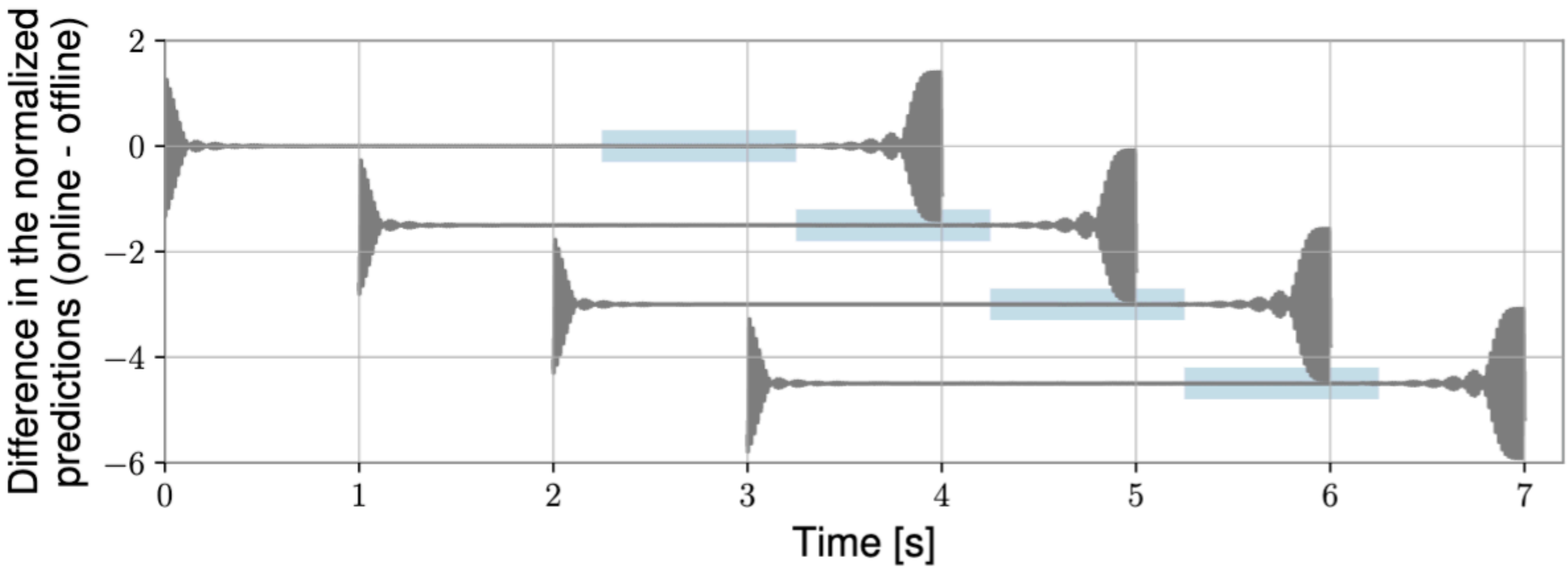
Ormiston et al (2020)

Online DeepClean: Specifications

- Aim is to subtract every 1 s frame as soon as they are available
- 1 s cleaned data are noisy at the edges
  - Reason: a combination of the architecture, training and post-processing (the major part).
- Consider 4 s data for cleaning and take 1 s from somewhere in the middle excluding the edges.
  - Causes additional latency

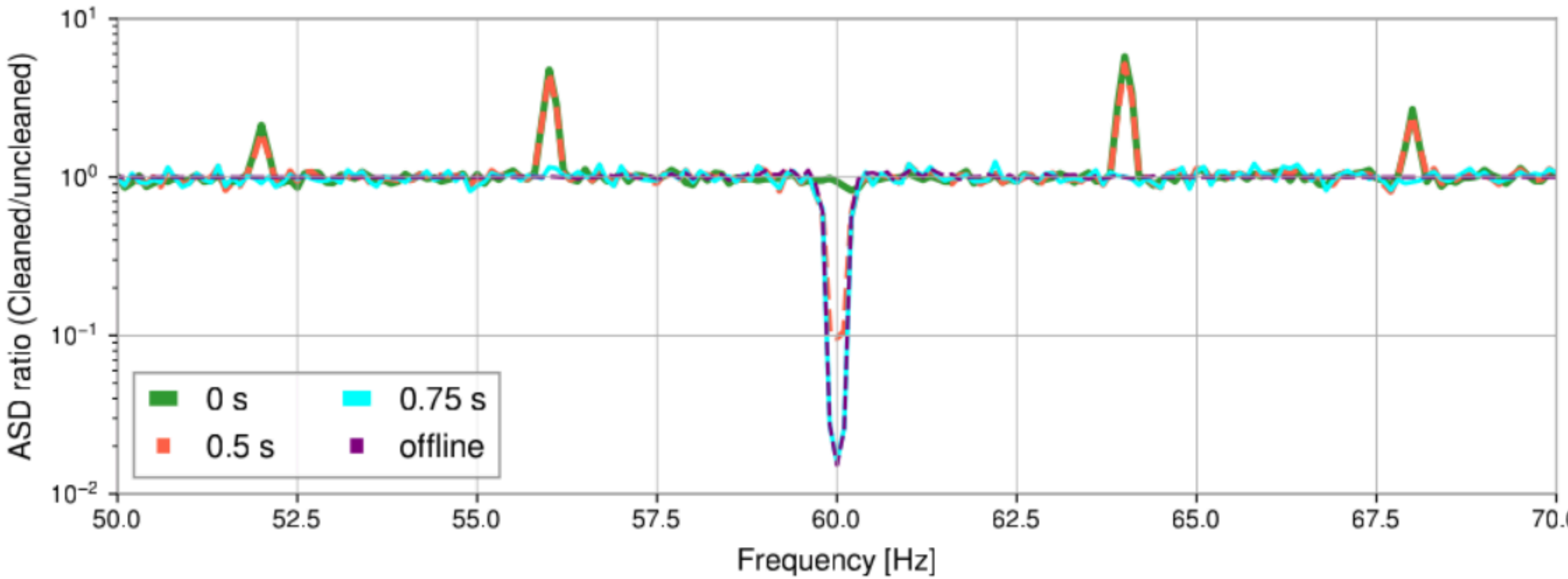
Will be discussed in other talks

Sliding Windows: Suppressing the Edge Effects



- Take unaffected 1s from the 4s cleaned segment and run cleaning on overlapping windows

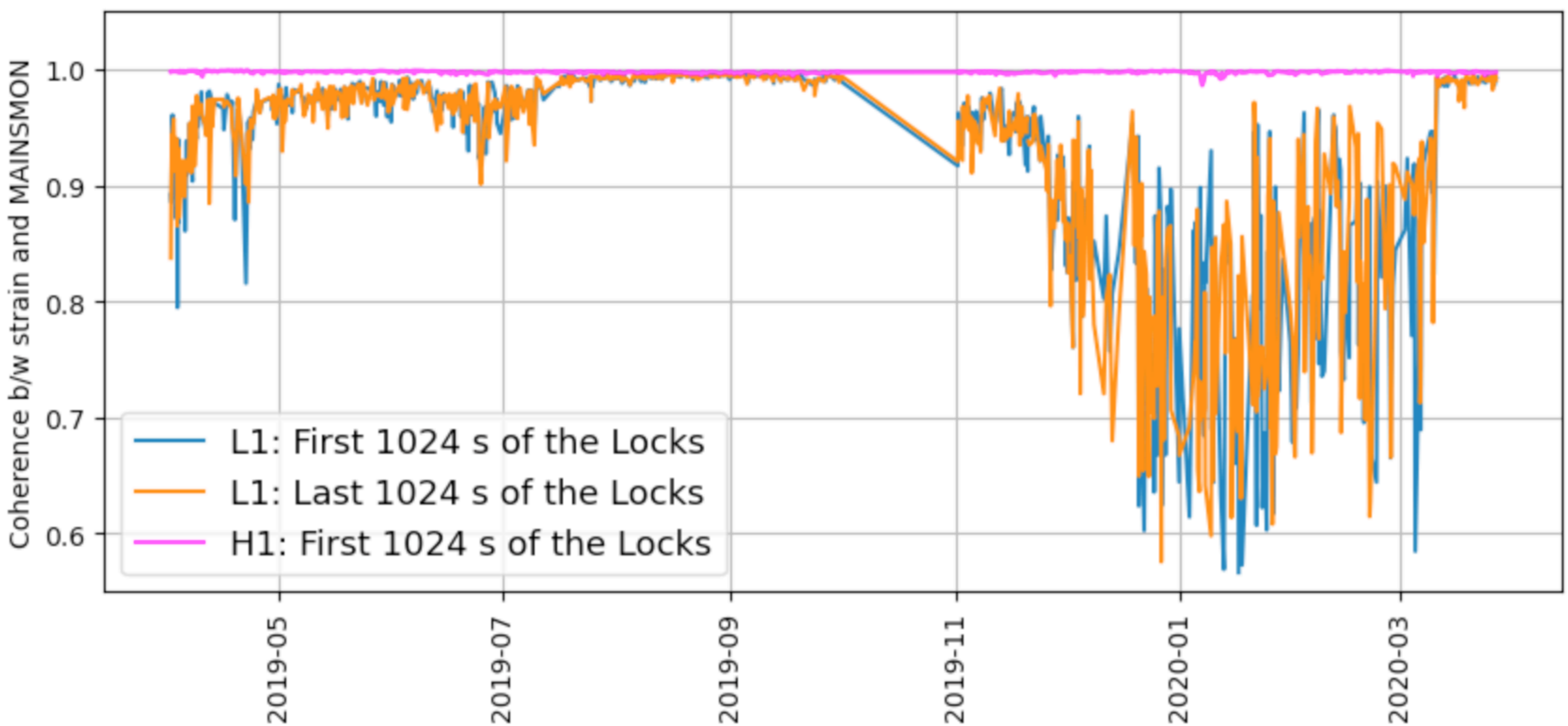
Online DeepClean: Latency vs Subtraction Quality



- Longer the excluded edge, better the quality of the subtraction

Gunny et al (2022)

COHERENCE BETWEEN THE STRAIN AND POWER-LINE WITNESS



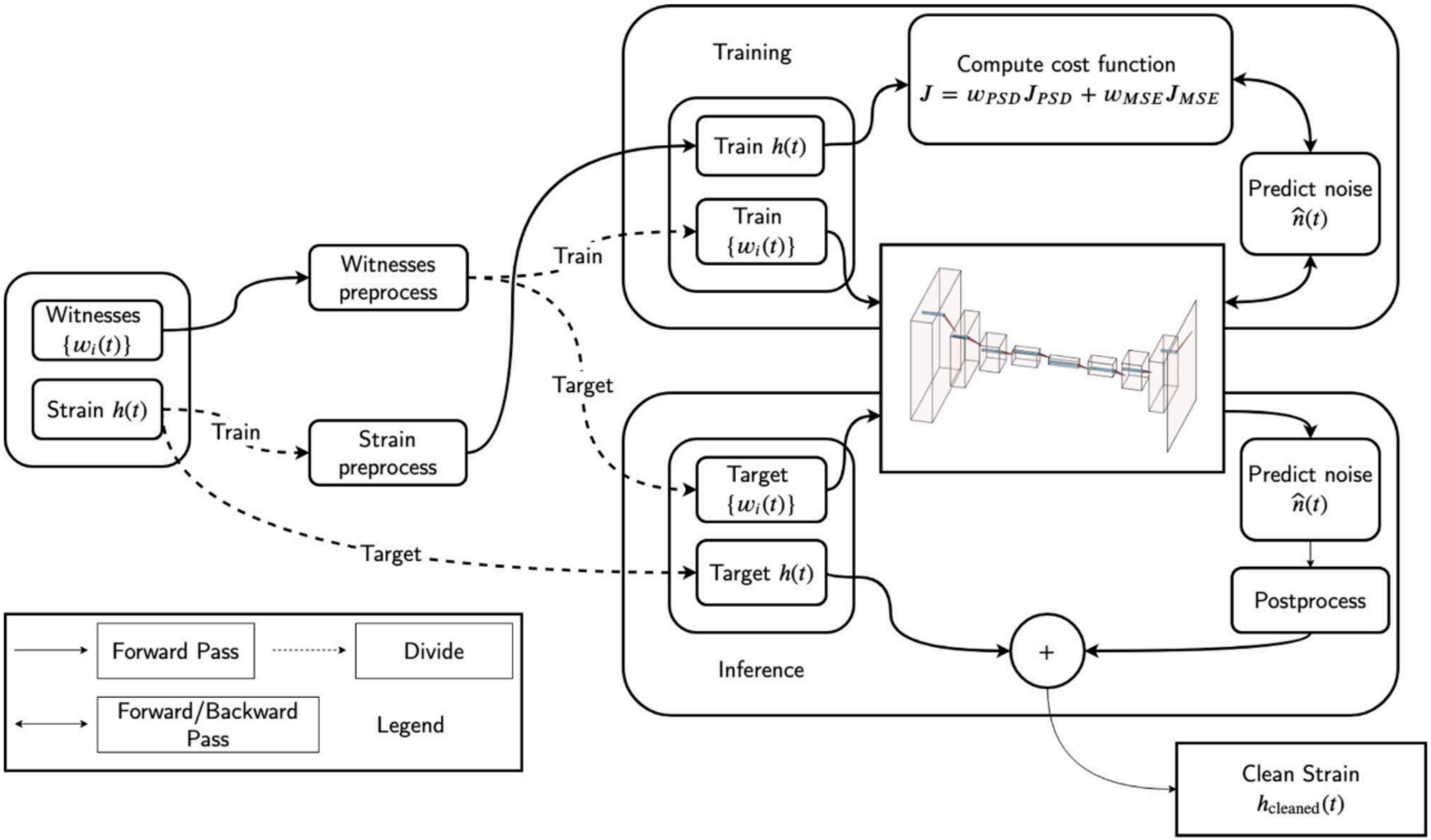
CONCLUSIONS

- DeepClean is able to subtract non-linear noise couplings in LIGO strain data that are witnessed by witness sensors
- Preparing for the fourth observing run as a Low latency pipeline
- Edges to be improved to further reduce latency
- While low coherence happens, need alternate approaches.
- Testing DeepClean on KAGRA data (talk by Chia-Jui)
- Deployment of computing infrastructure in proggress, that will help execution of fast ML algorithms (talk by Alec Gunny)

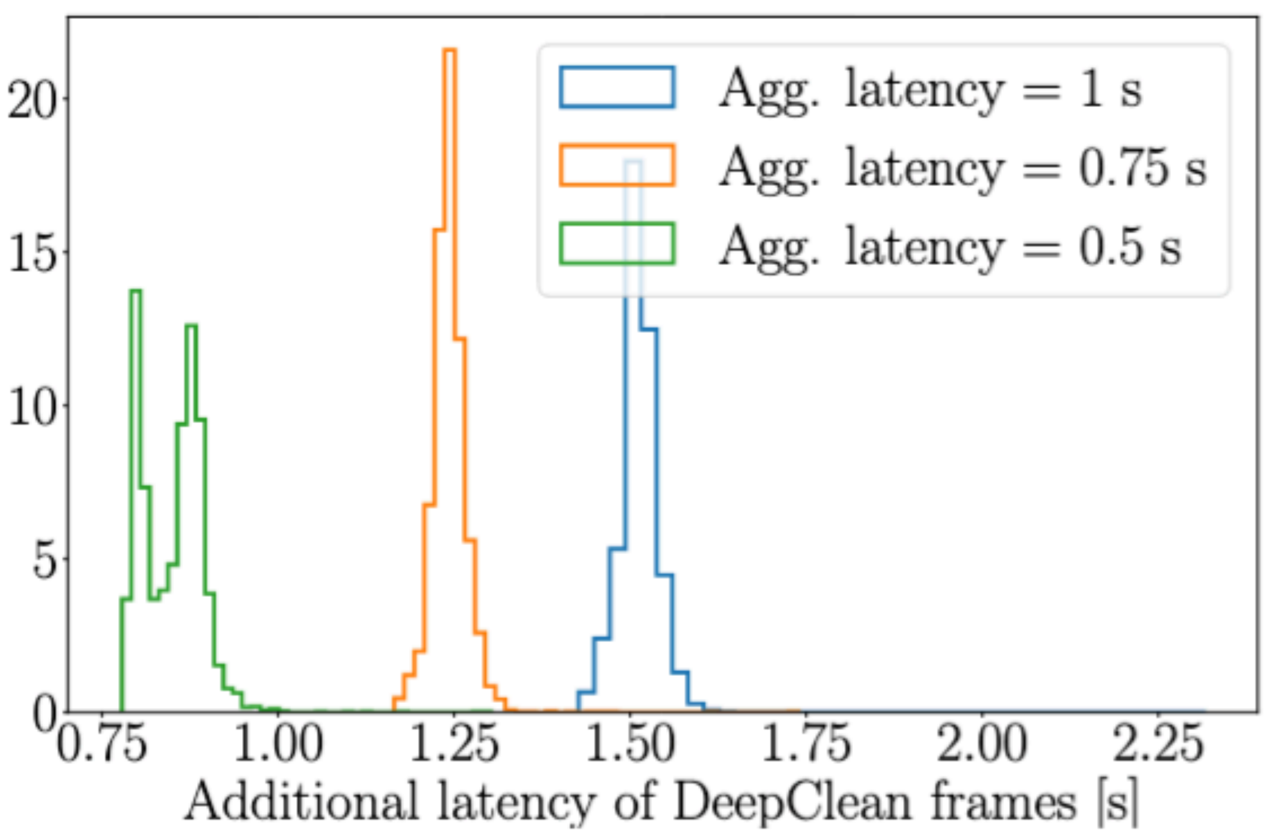
THANKS

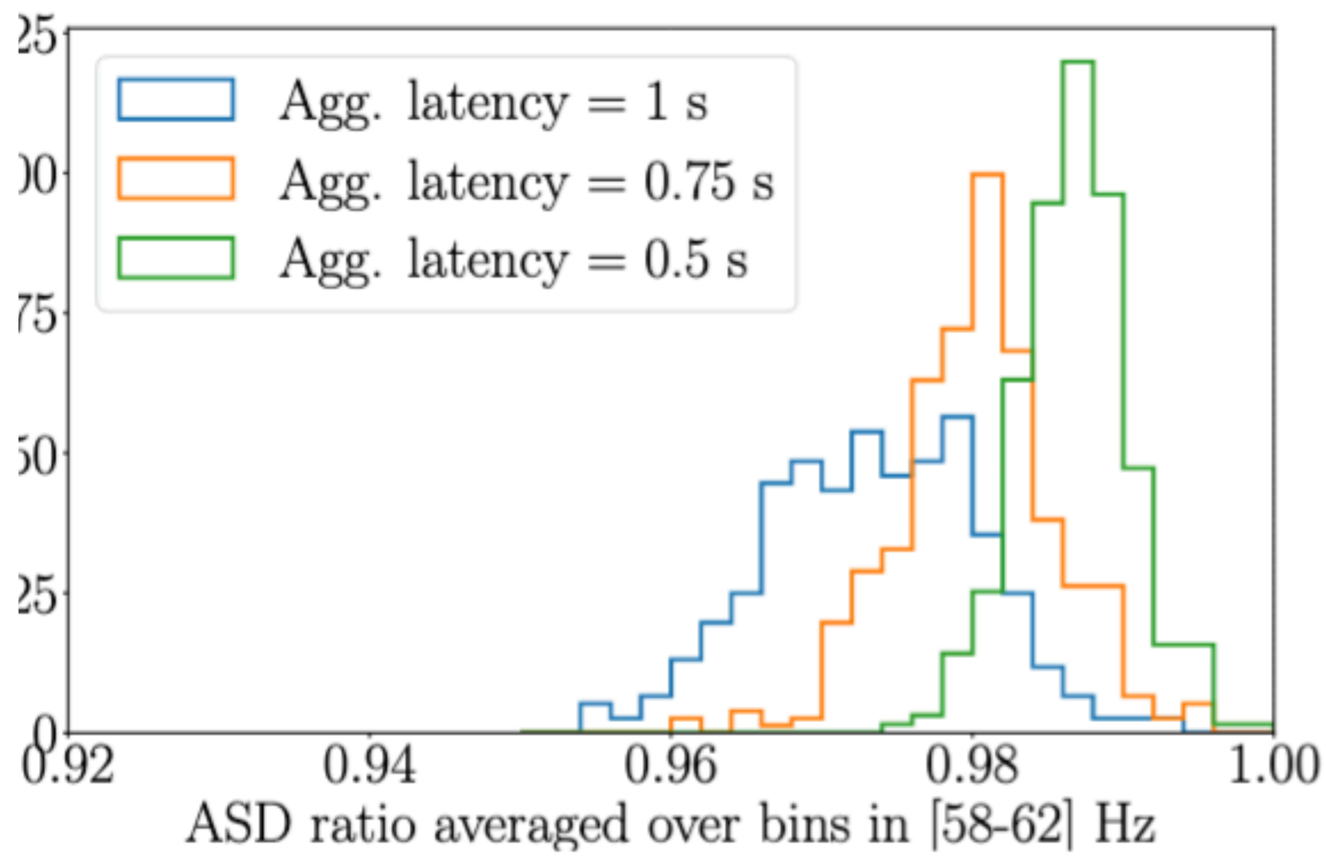
ADDITIONAL SLIDES

DEEPCLEAN: FULL WORKFLOW

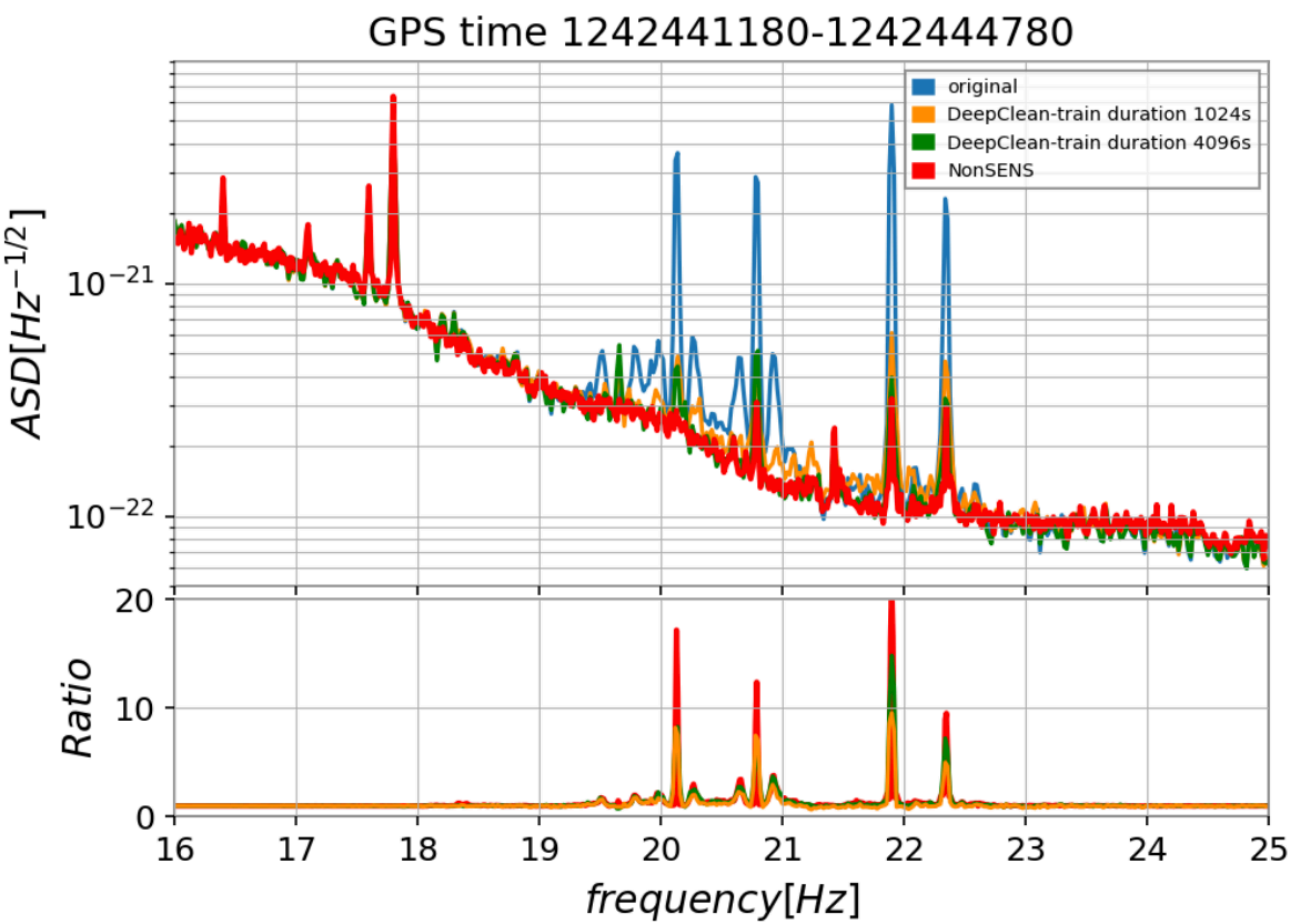


Ormiston et al (2020)


Agg. latency = 1 s
Agg. latency = 0.75 s
Agg. latency = 0.5 s
Additional latency of DeepClean frames [s]


Agg. latency = 1 s
Agg. latency = 0.75 s
Agg. latency = 0.5 s
ASD ratio averaged over bins in [58-62] Hz

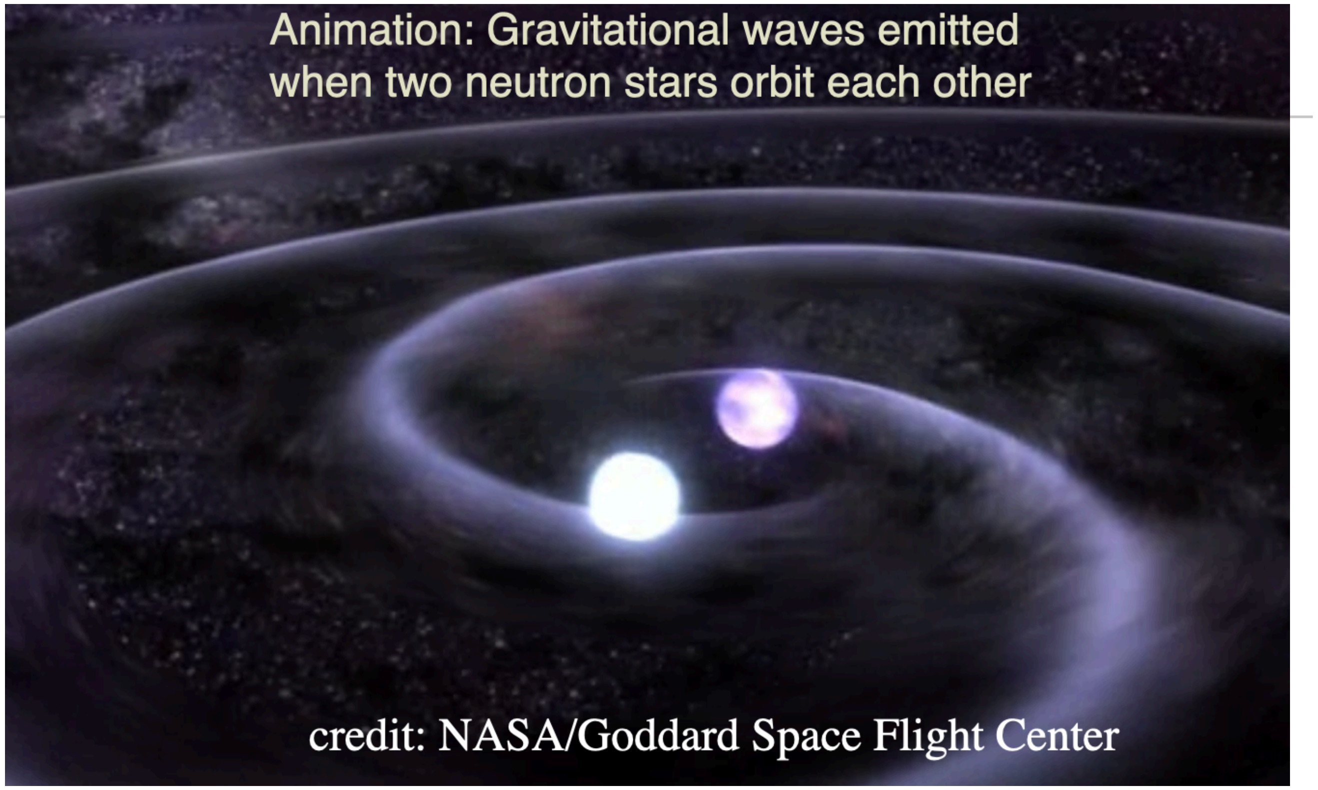
SUBTRACTION OF LOW-FREQUENCY BROADBAND NOISE (10-30 Hz)



Gravitational Waves (GW)

- Ripples in spacetime caused by massive objects moving with extreme acceleration (eg: pair of neutron stars and/or blackholes orbiting each other)

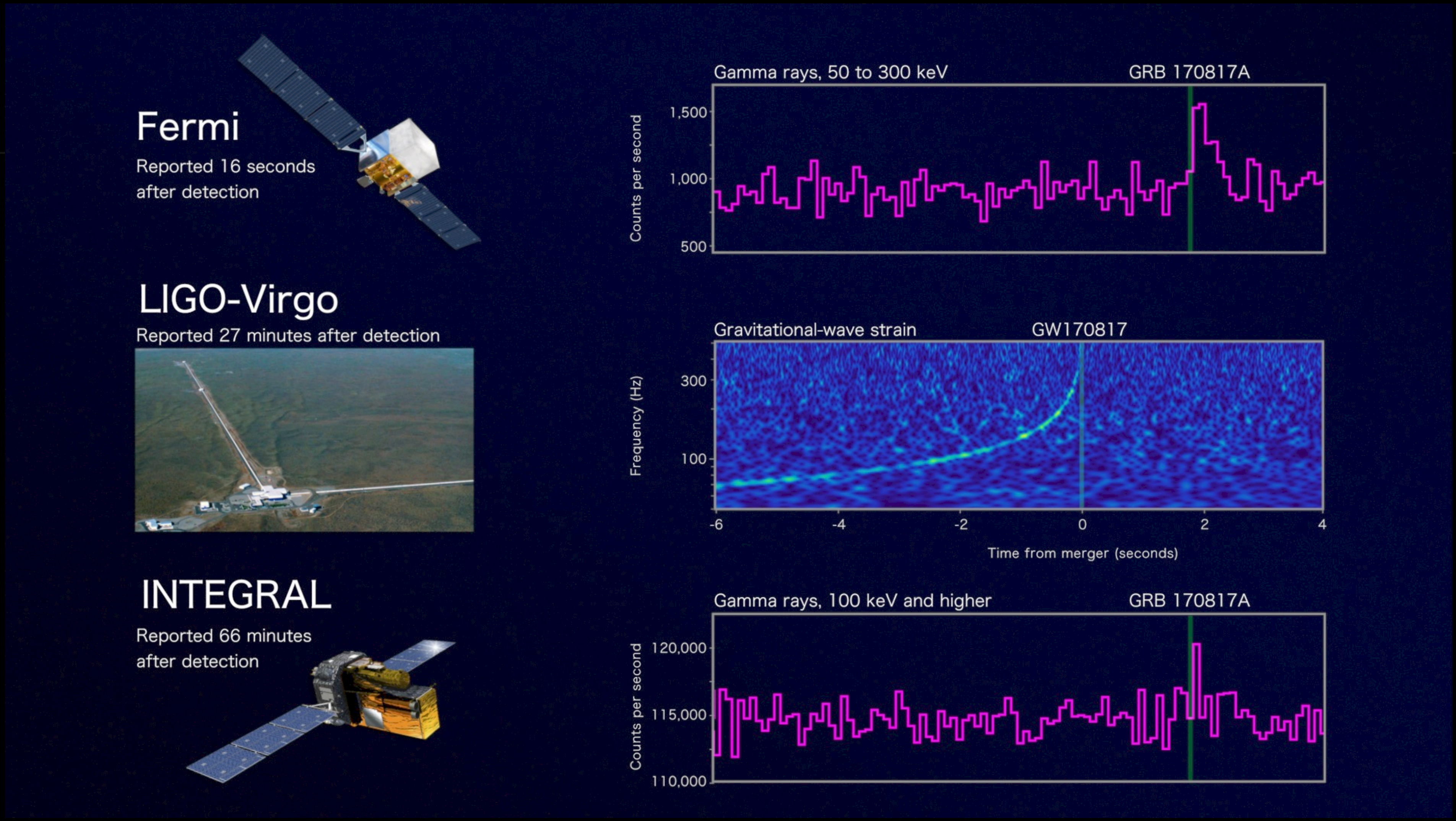
- Predicted by Albert Einstein in 1916 in his General theory of Relativity



- When they pass through earth, they produce strain $\sim 10^{-21}$ (producing displacements that are several thousands of times smaller than an atomic nucleus)

- Needs extremely sensitive precision instruments for detecting them.

MULTI-MESSENGER ASTRONOMY



- One event observed through multiple windows reveals complementary physics - step towards more complete picture of the universe
- GW detection alerts to be circulated at the earliest, to enable rapid follow up by EM telescopes in order to maximise the science outcome.

Offline DeepClean: Specifications

- Network trained with 2000 s (or 4000 s) data from around (up to a few days before or after) the target segment.
- Target segment is chosen to be 4096 s
- Batch: 32 overlapping kernels, with each kernel being 22 x 32768 matrices
    - Assuming there are 22 channels (for 60 Hz subtraction) and one kernel is 8 s and data is sampled at 4096 Hz
    - Batch size of 16 and kernels of 4s or 2s have been found to be equally effective in cleaning, however for training, the kernels need to be at least 8 s long as the loss function is PSD-based (a mean PSD with 4 FFTs and 1/2 Hz resolution needs 8s data)