Data Folding for the Stochastic Gravitational Wave Pipeline

Final Presentation
Eli Wiston
Mentors: Arianna Renzini and Colm Talbot

LIGO-T2100276-v1

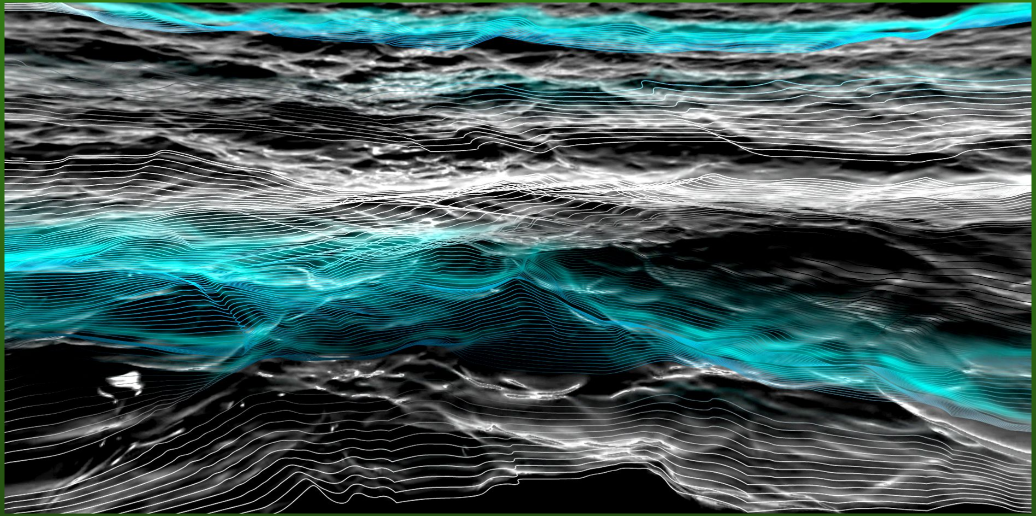
Outline

- Stochastic Signals
- Data folding: process and motivations
- Implementation in real data: O3a Run
- Results

Stochastic Gravitational Wave Background

- Sea of unresolved gravitational wave signals

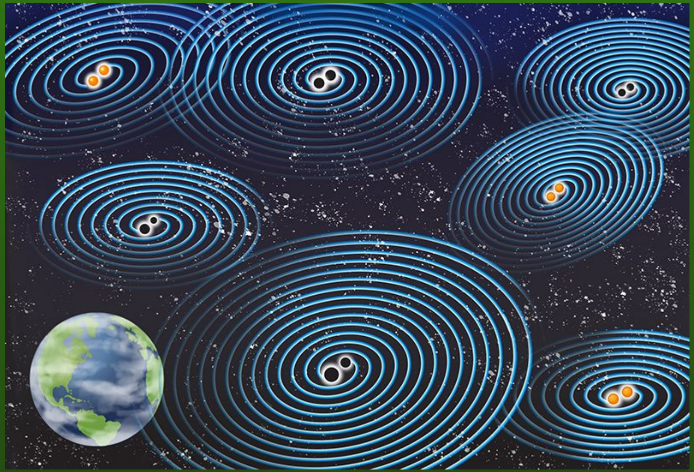
- Individually useless, collectively contain a lot of information



Why do we care?

There are two broad categories of stochastic signals:

- A low redshift, astrophysical foreground
  - Measure of large-scale structure independent of EM sources
- A high redshift, cosmological background
  - Evidence for inflation, early-universe phase transitions



How do we detect weak signals?

- Cross Correlation!
- The stochastic signal is coherent between the two detectors, while the noise is uncorrelated
- Over time, the noise will be suppressed relative to the signal
- But for such weak signals, we require long stretches of time

Stochastic Background Detection Statistic

Normalization

Assumed frequency distribution

Overlap function

Dirty Map

$$X = \frac{4}{\tau} \sum_{ft} \frac{H(f)\gamma^*_{ft,\alpha}}{P_{I_1}(t;f)P_{I_2}(t;f)} C_{1,2}(t;f)$$

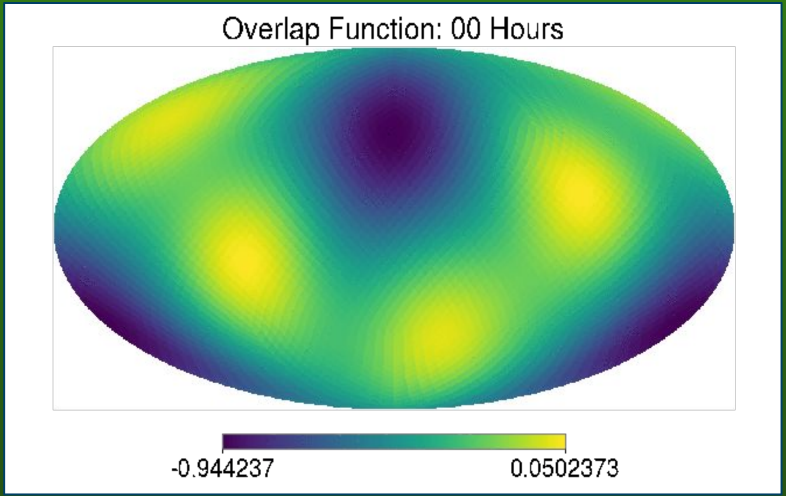
PSDs

CSD

Overlap Function

The overlap function is a measure of the sensitivity of a pair of detectors to different parts of the sky.

An anisotropic, stationary background will give the same signal at each part of each sidereal day.



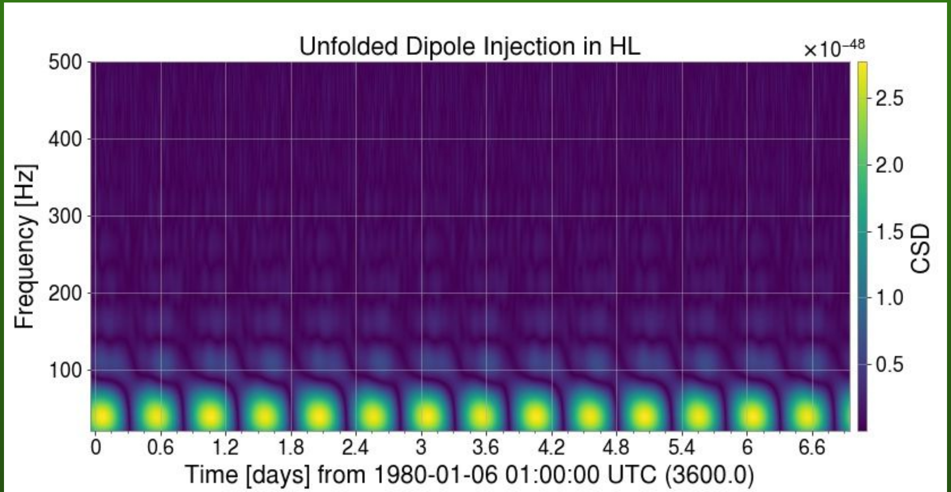
How does folding work?

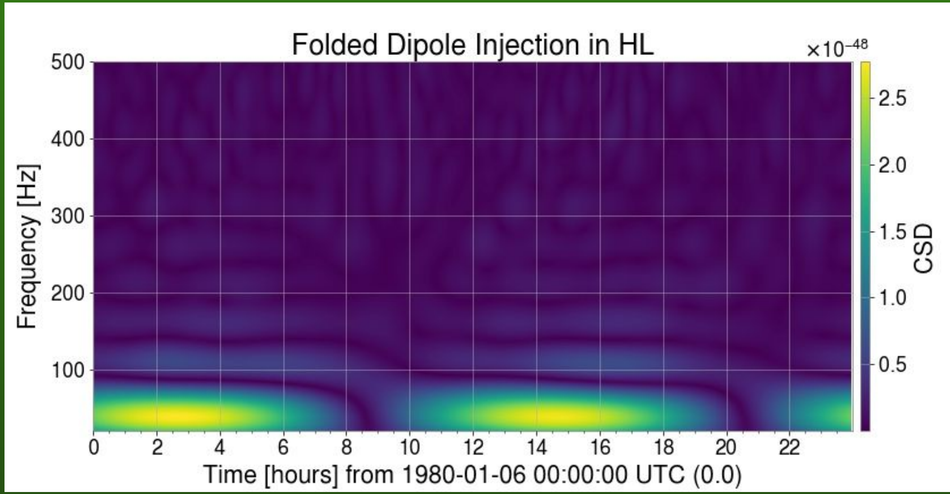
Exploiting the periodic nature of the overlap function, we can re-write these equations as:

$$X = \frac{4}{\tau} \sum_f H(f) \sum_{ft_s} \gamma^{I*}_{ft_s,\alpha} \sum_{i_{\text{day}}} \frac{CSD(i_{\text{day}} T_s + t_s; f)}{PSDs(i_{\text{day}} T_s + t_s; f)}$$

This sum over i_{day} is our data folding. This sum only needs to be performed once, but will speed up subsequent analysis by a factor of ~N_{day}.

Folding on Simulated Data



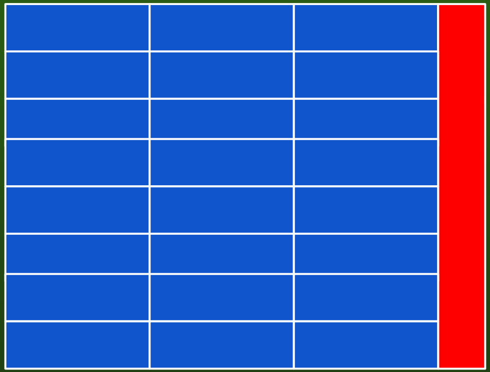


Mitigating Information Loss in Folding

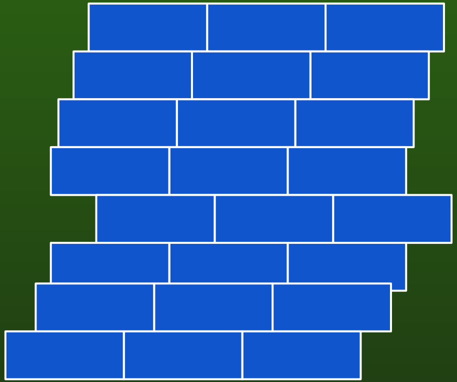
Unfolded method:

Problem: fft lengths do not always fit evenly into one sidereal day

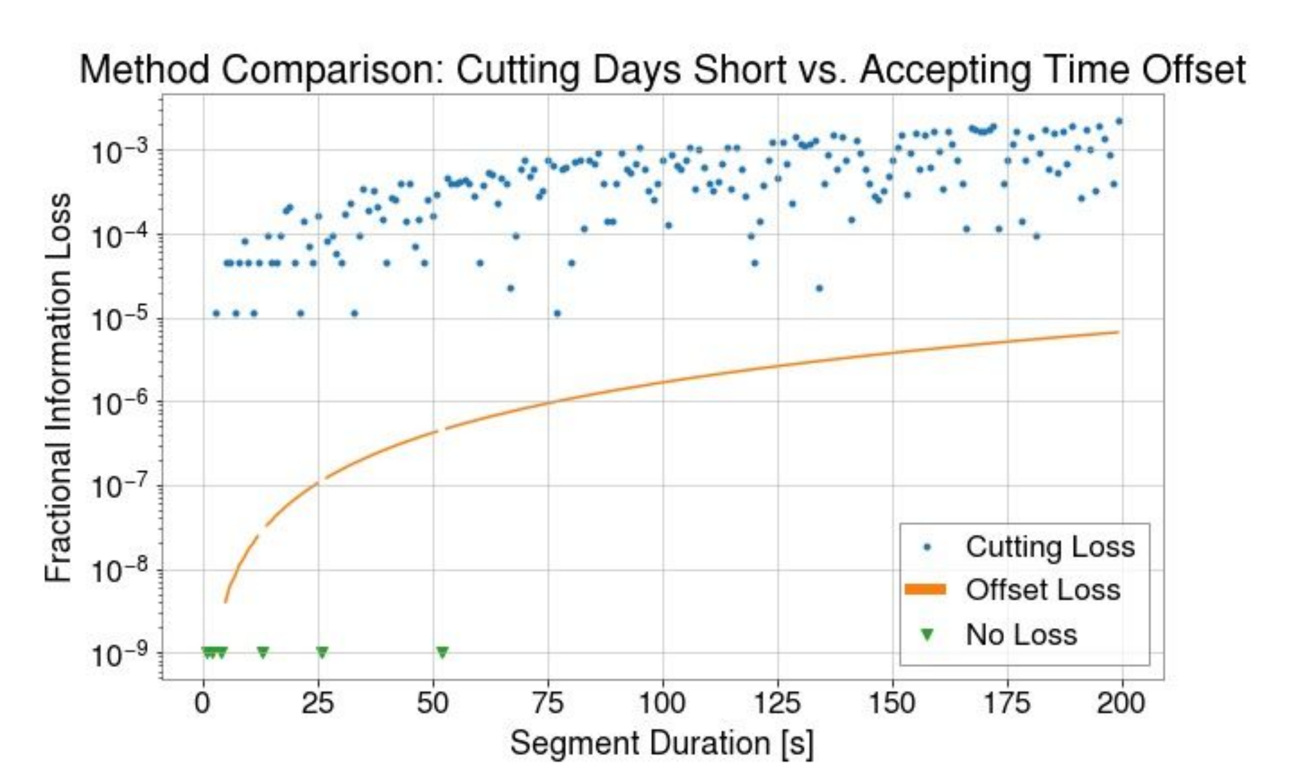
Method 1: Cut days short



Method 2: Keep all times with some offset



Mitigating Information Loss in Folding

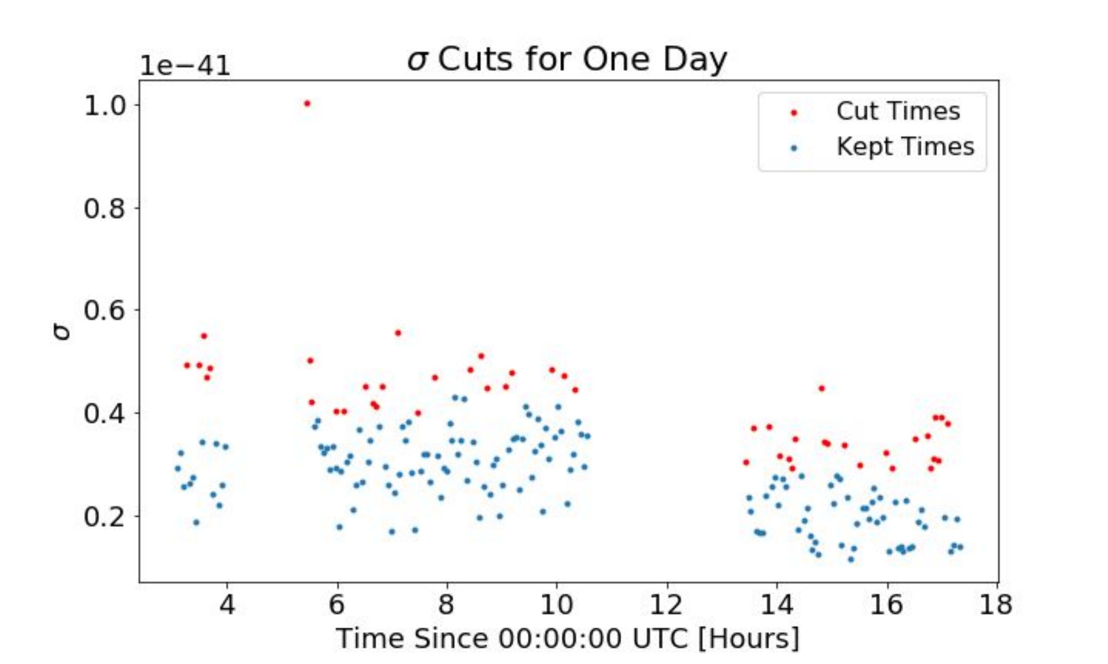


Data Quality - Delta Sigma Cuts

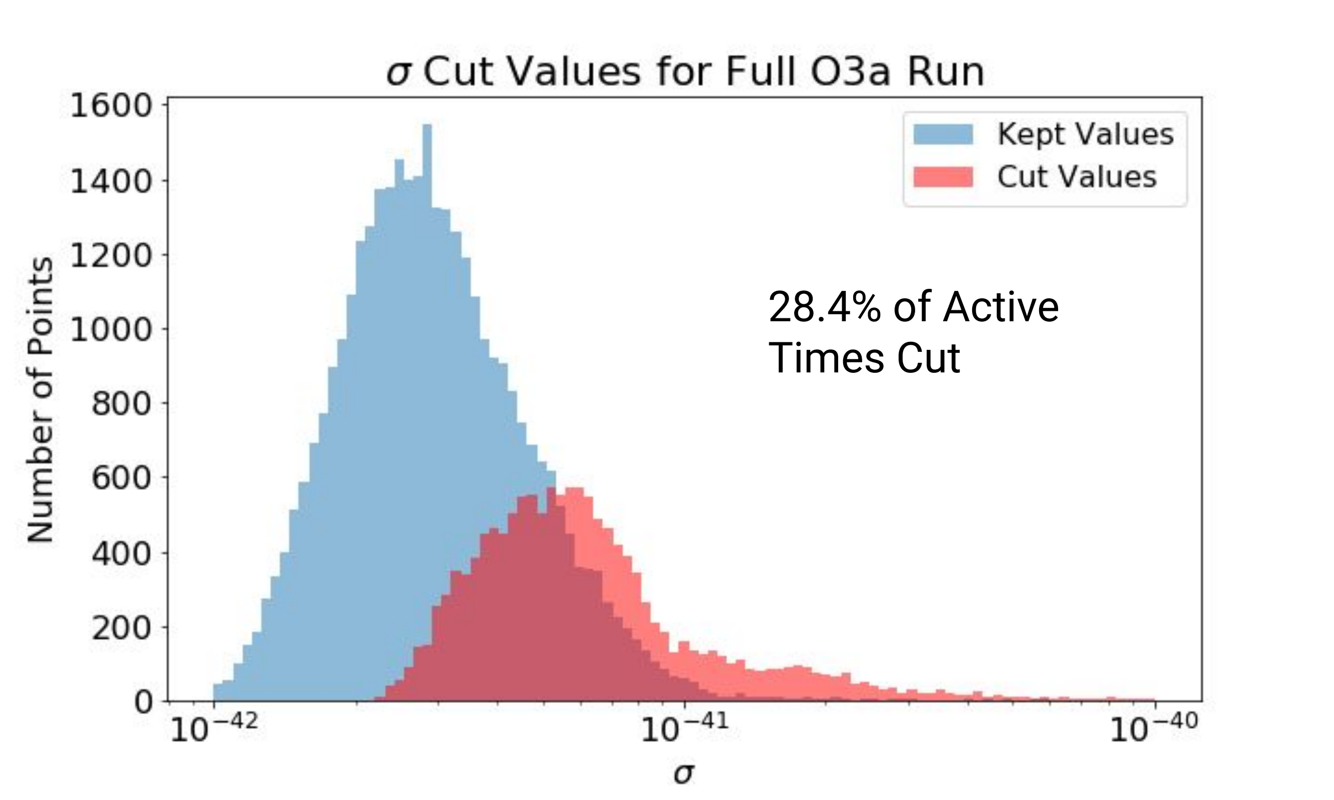
$$\sigma = \sqrt{\sum_f \left(H(f) PSD_1 PSD_2\right)}$$

For H(f) = 1:

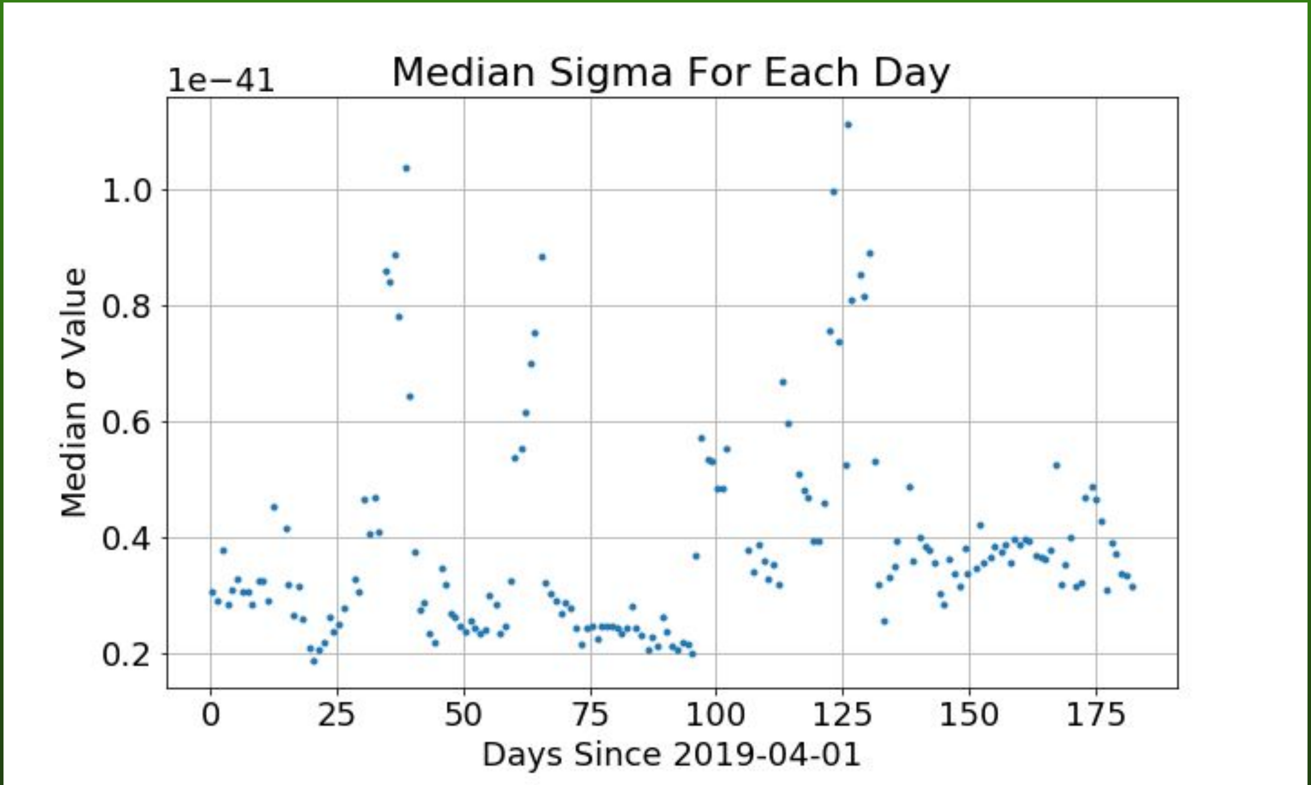
$$\sigma = \sqrt{\sum_f (PSD_1 PSD_2)}$$



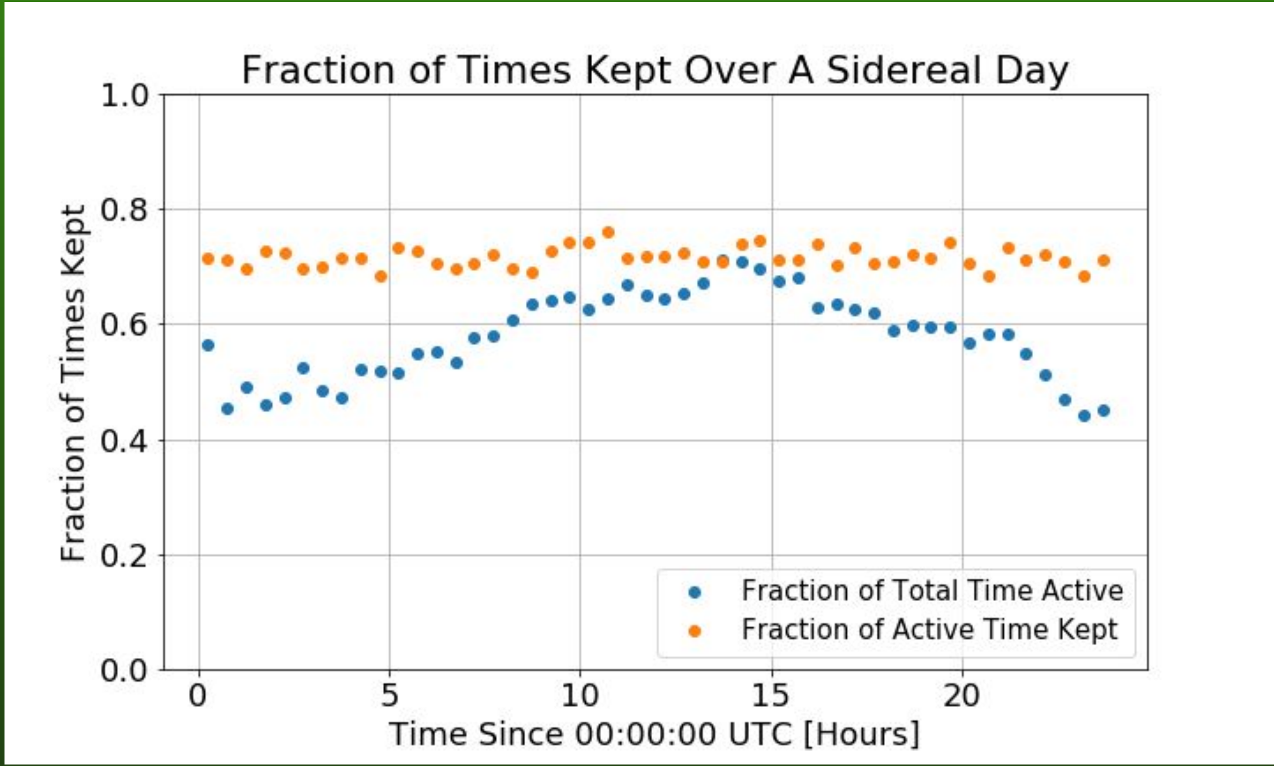
Sigma Cuts -Full O3a Dataset



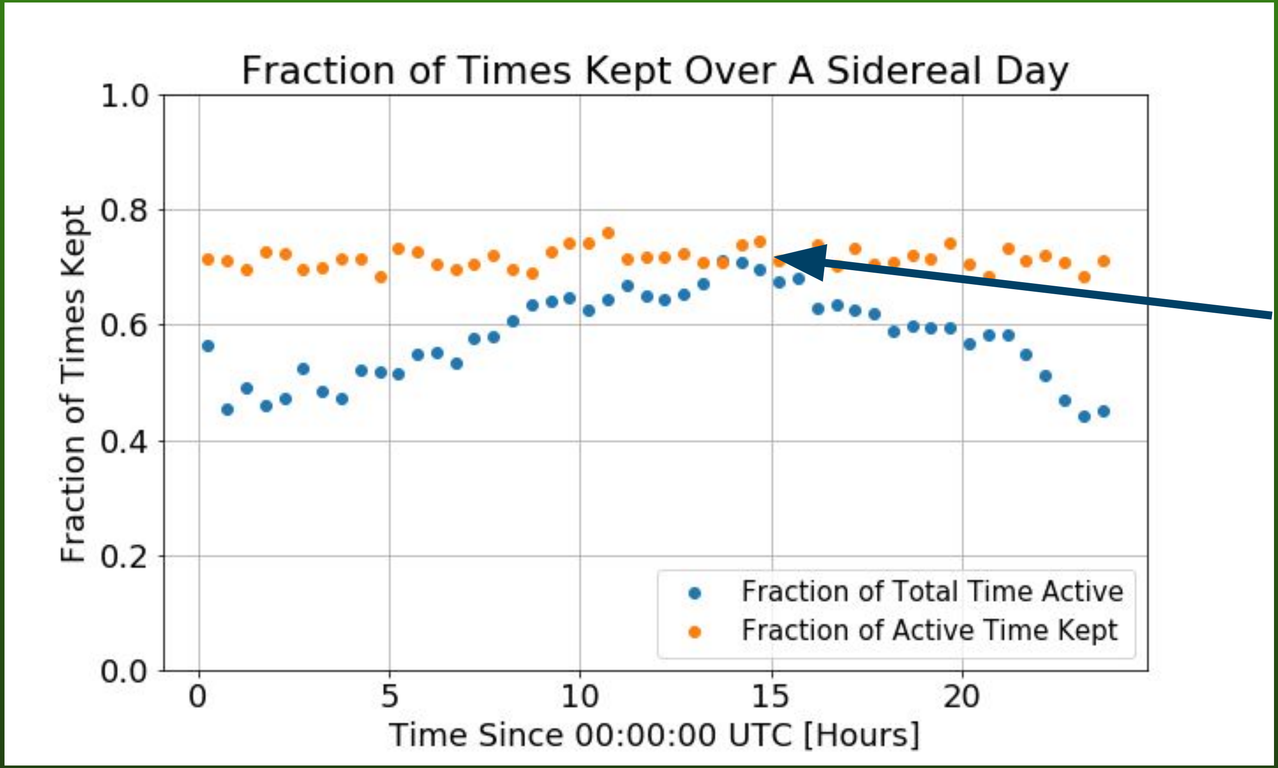
Sigma Variance Across the Run



Sampling Variance Across A Sidereal Day



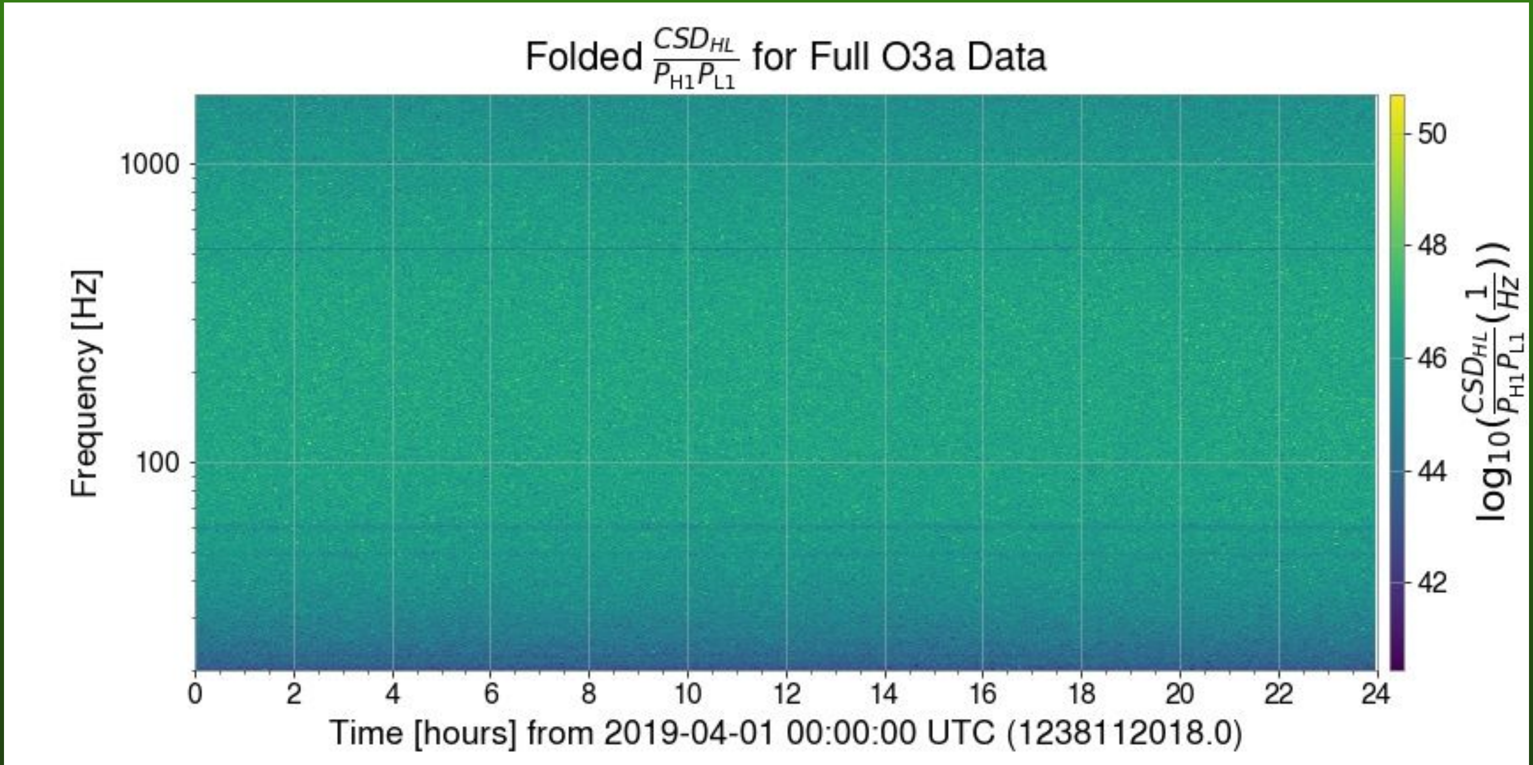
Sampling Variance Across A Sidereal Day



7:00 AM in H1

9:00 AM in L1

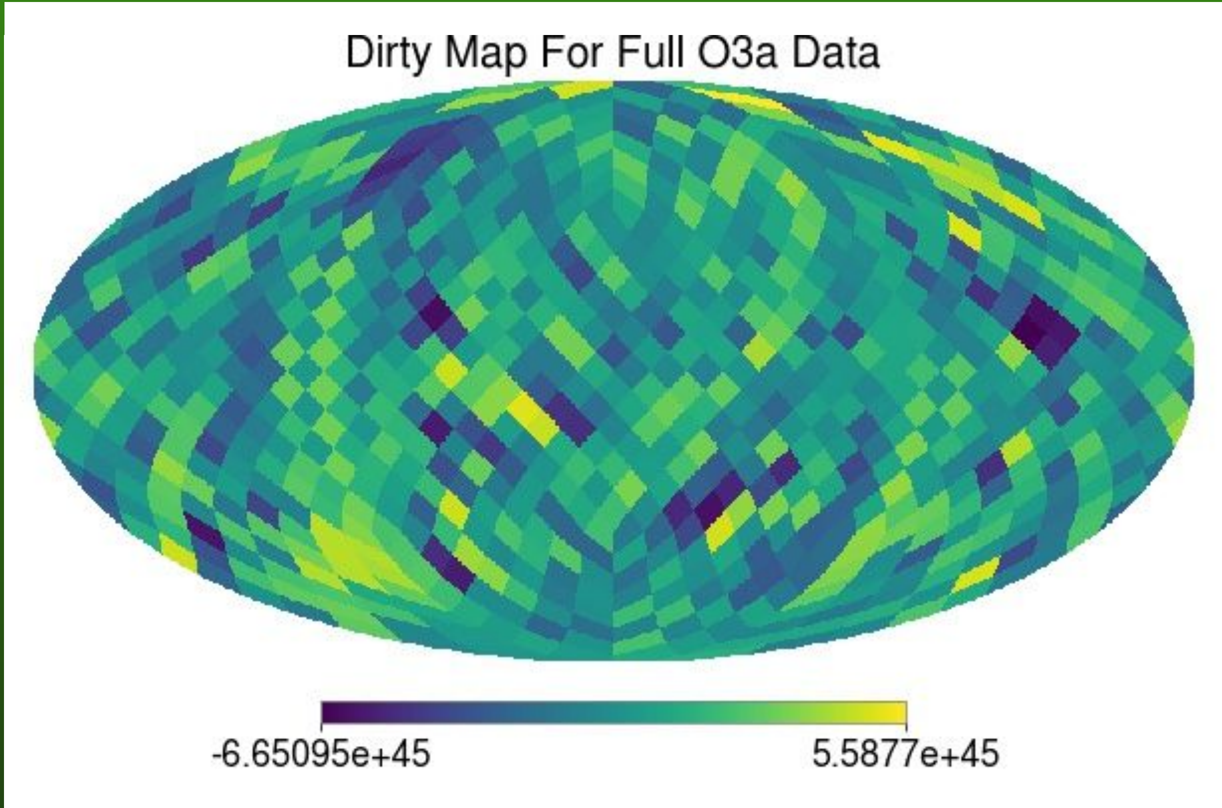
Final Folded Spectrogram



Dirty Map

Computation time on folded data: 11 hours

Estimated computation time for unfolded: 84 days



Summary

- Stochastic data can provide a range of astrophysical and cosmological insights
- The periodic nature of the overlap function and assumed stationarity of the signal allow us to average the data down to the size of a sidereal day
- Folding CSD/PSDs allows for efficient computation of the dirty map, our key stochastic background detection statistic

Acknowledgements

I'd like to thank:

My mentors, Arianna Renzini and Colm Talbot

Derek Davis

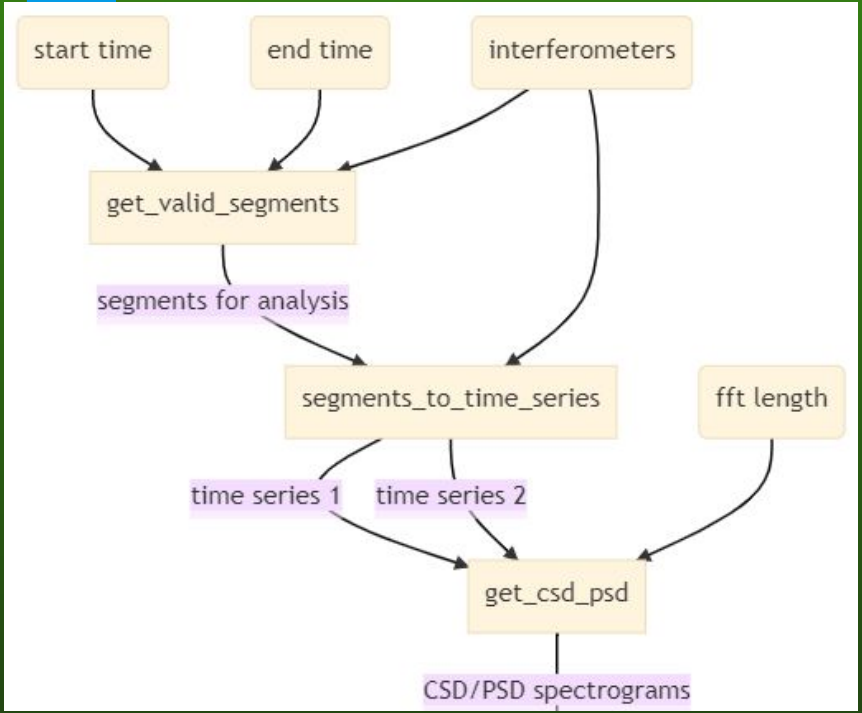
Alan Weinstein

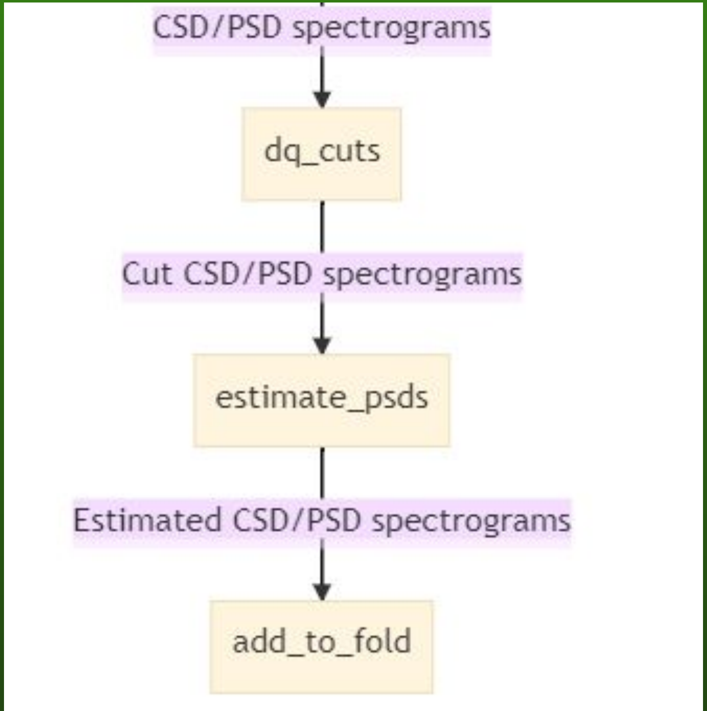
My fellow LIGO SURFs

The NSF and LIGO Lab

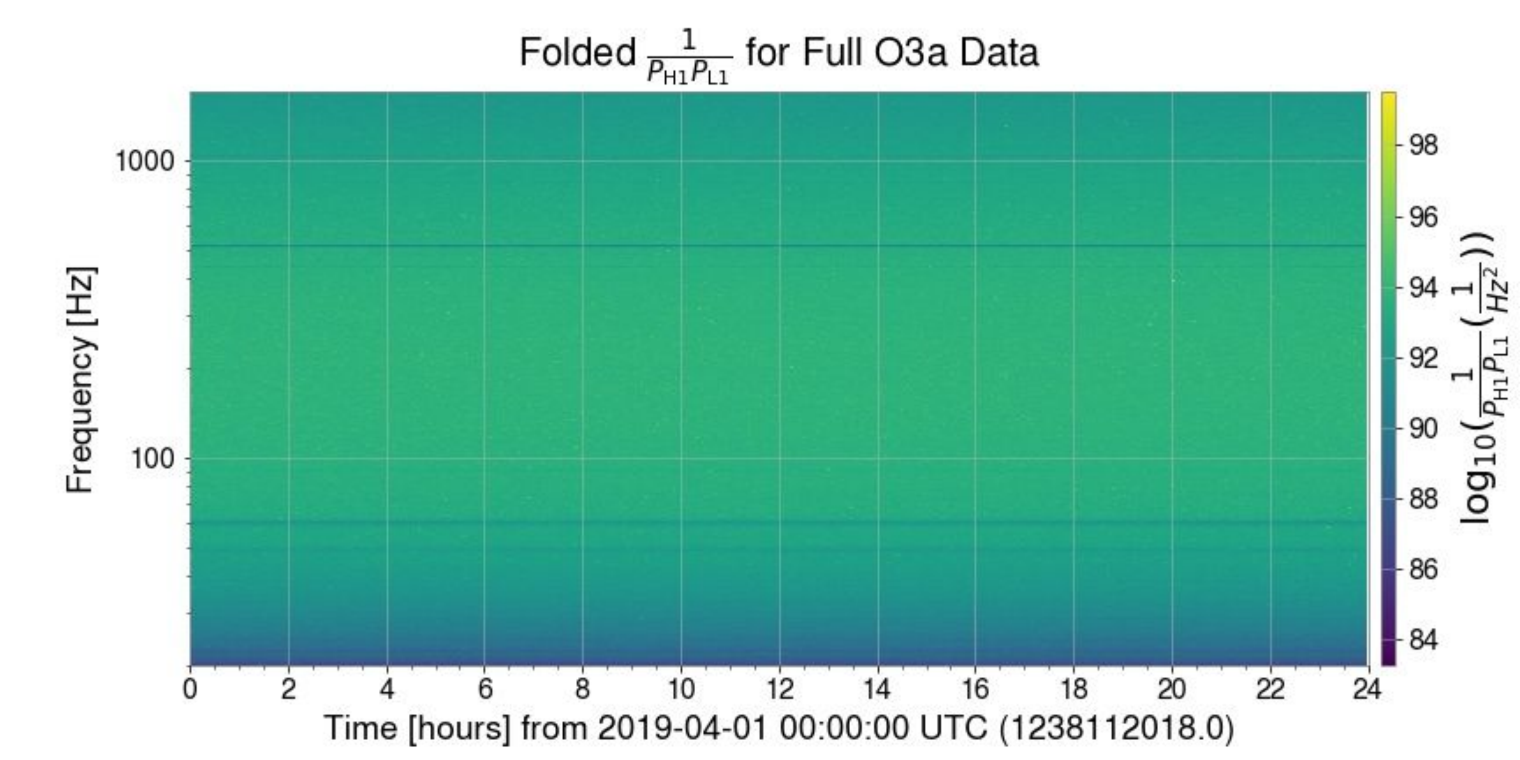
Backup Slides

Workflow


start time
end time
interferometers
get_valid_segments
segments for analysis
segments_to_time_series
fft length
time series 1
time series 2
get_csd_psd
CSD/PSD spectrograms


CSD/PSD spectrograms
dq_cuts
Cut CSD/PSD spectrograms
estimate_psds
Estimated CSD/PSD spectrograms
add_to_fold

Final Spectrograms



Fisher Information Matrix

$$\Gamma = 4 \sum_{Ift} \frac{H^2(f)}{P_{I_1}(t;f)P_{I_2}(t;f)} \gamma^{I*}_{ft,\alpha} \gamma^{I}_{ft,\alpha'}$$