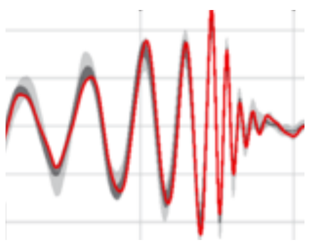
Rich Ormiston

MEASURING THE STOCHASTIC GRAVITATIONAL WAVE BACKGROUND WITH LIGO

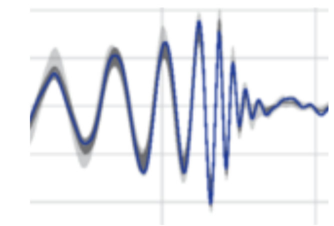
January 31, 2019

Overview

- Description of gravitational waves
- What is the stochastic gravitational wave background (SGWB) and why do we care?
- How can we test for the SGWB with LIGO and what are the current results?
- Can we enhance the sensitivity of the SGWB search through nonlinear data cleaning via machine learning?



Gravitational Waves



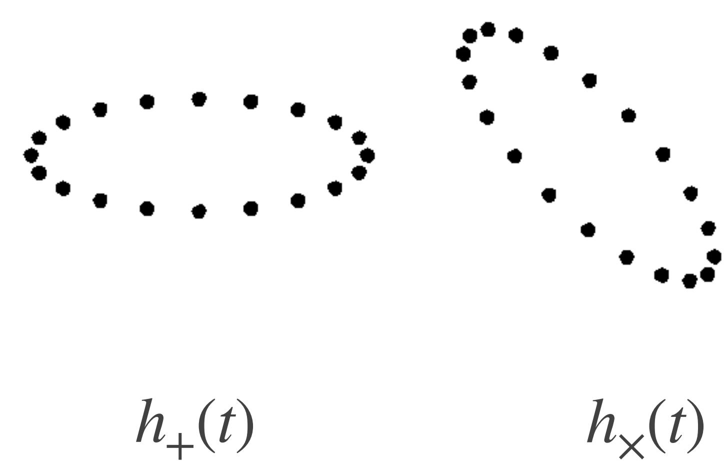
- Gravitational waves are first order perturbations to the spacetime metric.
- GWs were theorized by Poincare in 1905, formalized by Einstein in 1915, and only *indirectly* measured for the first time in 1974 ([Hulse-Taylor binary]).
- The first direct detection occurred on September 14, 2015 by the LIGO & Virgo Collaboration ([GW150914])

$$\frac{\Delta L}{L} = \frac{1}{2}h(t) = \frac{1}{2}\left(h_+\hat{e}^{+}_{ab} + h_\times\hat{e}^{\times}_{ab}\right)D^{ab}$$

GW Strain

Detector Tensor

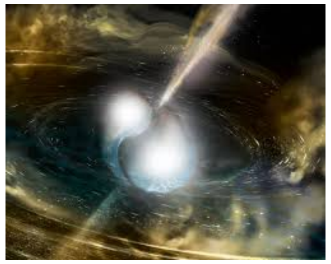
Fractional change in IFO arm length



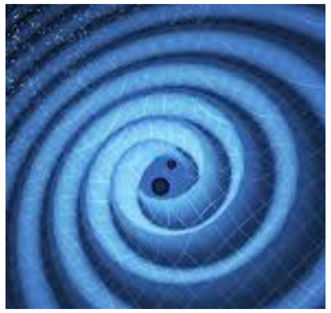
Effects of "plus" (left) and "cross" (right) gravitational wave polarizations on a ring of test masses

Sources of Gravitational Waves

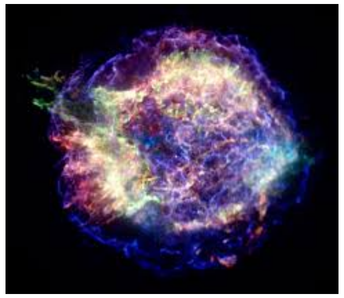
Binary Black Holes



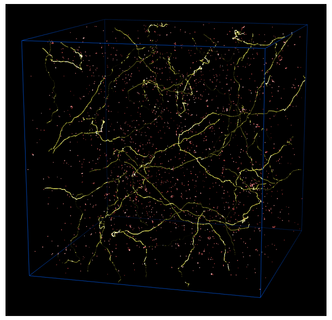
Binary Neutron Stars



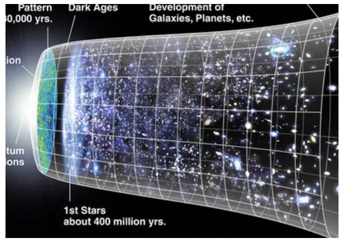
Black Hole +
Neutron Star



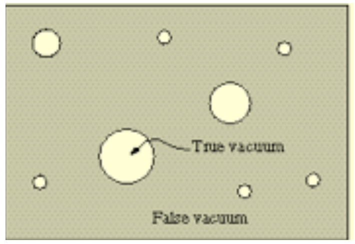
Supernovae



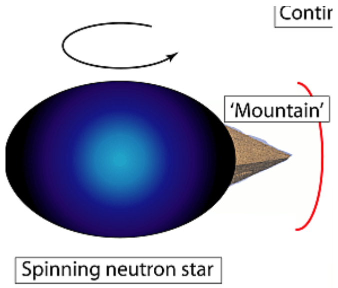
Cosmic Strings



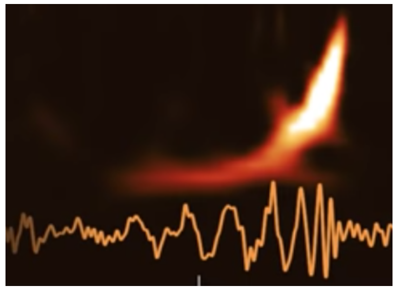
Inflation



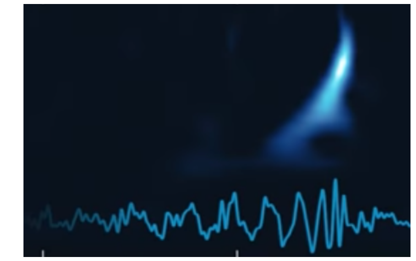
Phase Transitions



Non-symmetric
Neutron Stars



From GR to GW



We start by expanding Einstein's equation to first order in the metric. Assuming $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and $|h_{\mu\nu}| \ll 1$, we have

$$\frac{8\pi G}{c^4} T_{\mu\nu} = R^{(1)}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R^{(1)}$$

$$= \frac{1}{2}\left(-\frac{\partial^2 h}{\partial x^\mu \partial x^\nu} + \frac{\partial^2 h^\alpha_\nu}{\partial x^\mu \partial x^\alpha} + \frac{\partial^2 h^\alpha_\mu}{\partial x^\nu \partial x^\alpha} - \eta^{\alpha\beta} \frac{\partial^2 h_{\mu\nu}}{\partial x^\alpha \partial x^\beta} \right) - \frac{1}{2}\left(\frac{\partial^2 h^{\mu\nu}}{\partial x^\mu \partial x^\nu} - \eta^{\mu\nu} \frac{\partial^2 h}{\partial x^\mu \partial x^\nu} \right)$$

Adopting the transverse traceless gauge simplifies the expression greatly

$$\boxed{-\frac{16\pi G}{c^4} T_{\mu\nu} = \Box \bar{h}_{\mu\nu}}$$

Gravitational Waves!

The Stochastic GW Background I

The SGWB is the superposition of many sources emitting gravitational waves which are too weak to be individually detected.

Normalized Energy Density from GWs → $\Omega_{GW}(f) = \frac{1}{\rho_{c,0}} \frac{d\rho_{GW}(f)}{d\log f}$ ← Energy density from GW contained in interval $f \to f + \Delta f$

Energy Density Required to Close the Universe $= 3H_0^2c^2/(8\pi G)$

The Stochastic GW Background I

The SGWB is the superposition of many sources emitting gravitational waves which are too weak to be individually detected.

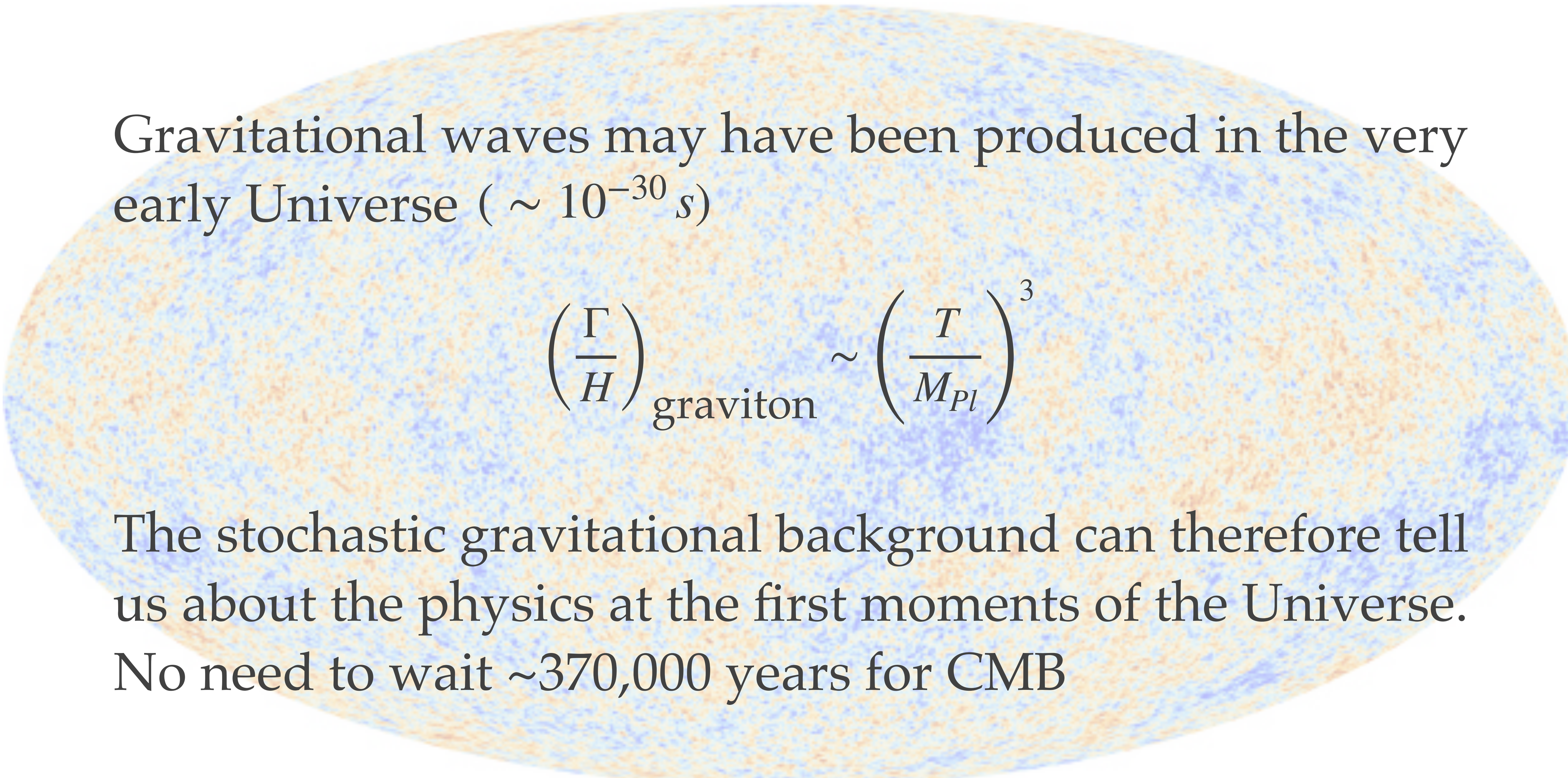
$$h_{100}^2\,\Omega_{GW}(f) = \frac{h_{100}^2}{\rho_{c,0}}\frac{d\rho_{GW}(f)}{d\log f}$$

$$H_0 = h_{100} \times 100 \text{ km s}^{-1}\text{ Mpc}^{-1} = 67.9 \text{ km s}^{-1}\text{ Mpc}^{-1}$$

Remove our ignorance of the Hubble constant from the analysis

Ref: Abbot et. al July 2017

The Stochastic GW Background II



Gravitational waves may have been produced in the very early Universe ($\sim 10^{-30}\,s$)

$$\left(\frac{\Gamma}{H}\right)_{\text{graviton}} \sim \left(\frac{T}{M_{Pl}}\right)^3$$

The stochastic gravitational background can therefore tell us about the physics at the first moments of the Universe. No need to wait ~370,000 years for CMB

Ref: [Planck 2013]

The Stochastic GW Background III

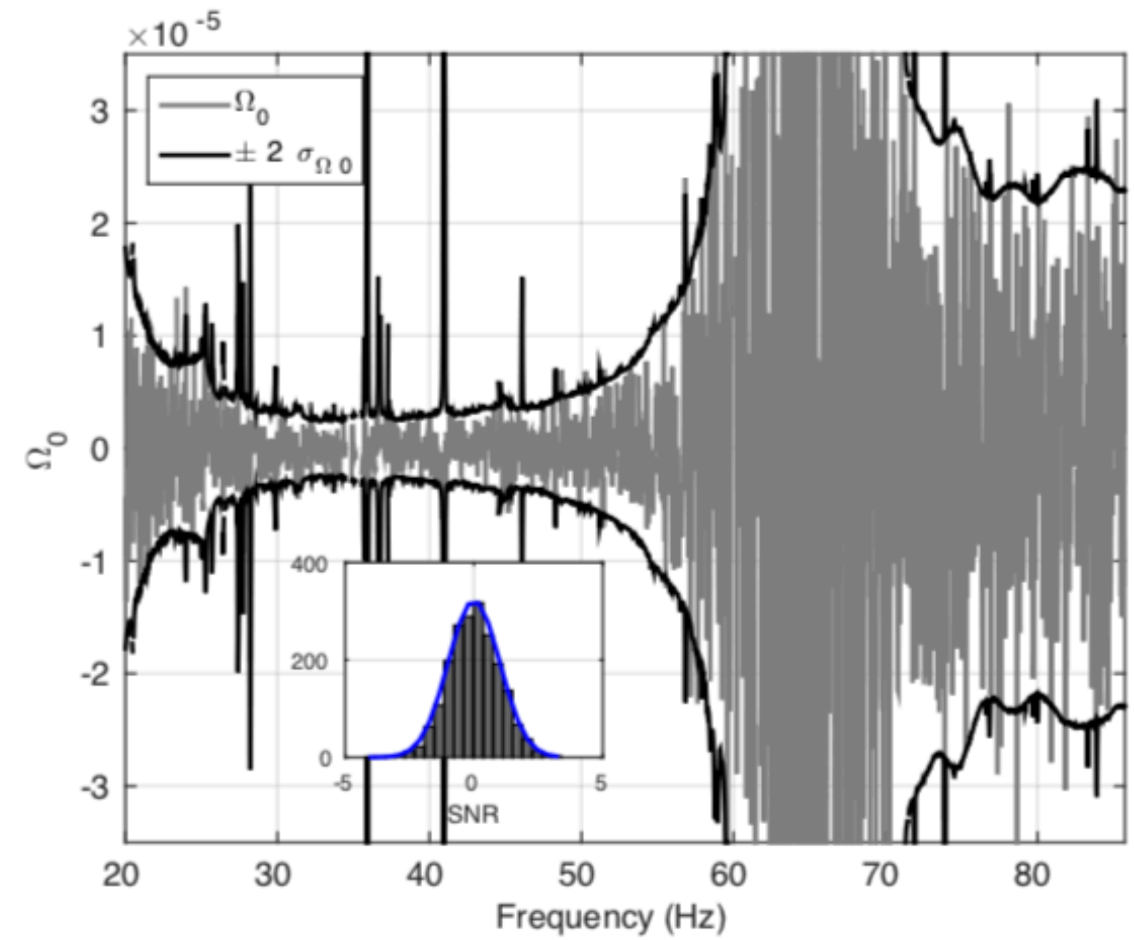
Assumptions about SGWB:

- Isotropic
- Gaussian
- Unpolarized
- Stationary

Because random noise is also gaussian (and loud), we cannot use a single detector to measure the background. Instead, we must *cross-correlate* the signal S between two detectors:

$$S = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{s}_H^*(f)\tilde{s}_L(f')\tilde{Q}(f')\delta(f-f')\,df\,df'$$

$$\tilde{s}_i(f) = \tilde{n}_i(f) + \tilde{h}_i(f)$$



Bin-by-bin estimator for O1 isotropic search with uncertainly. Inset plot is SNR over frequency bins with gaussian fit

Ref: Abbot et. al July 2017

Calculating the SGWB Energy Density

To find $\rho_{GW}(f)$ we consider the T_{00} component of the stress energy tensor

$$\rho_{GW}(f) = T_{00} = \frac{c^2}{32\pi G}\left\langle \dot{h}^{ij}_{TT}\dot{h}^{TT}_{ij}\right\rangle$$

We then transform to frequency space (makes life easier) and can write the 2-point correlation

$$\left\langle h^*_A(f,\hat{\Omega})h_{A'}(f',\hat{\Omega}')\right\rangle = \delta^2(\hat{\Omega},\hat{\Omega}')\delta(f-f')\delta_{AA'}\frac{1}{16\pi}H(f)$$

Using the plane-wave expansion for the GWs and plugging it all in we get

$$\int df\,\frac{d\rho_{GW}}{df} = \frac{\pi c^2}{4G}\int_0^\infty df\, f^2 H(f) = \frac{\pi c^2}{4G}\int_0^\infty df\, f^2 \frac{3H_0^2}{2\pi^2}\frac{\Omega_{GW}(f)}{f^3}$$

Calculating the SGWB Energy Density

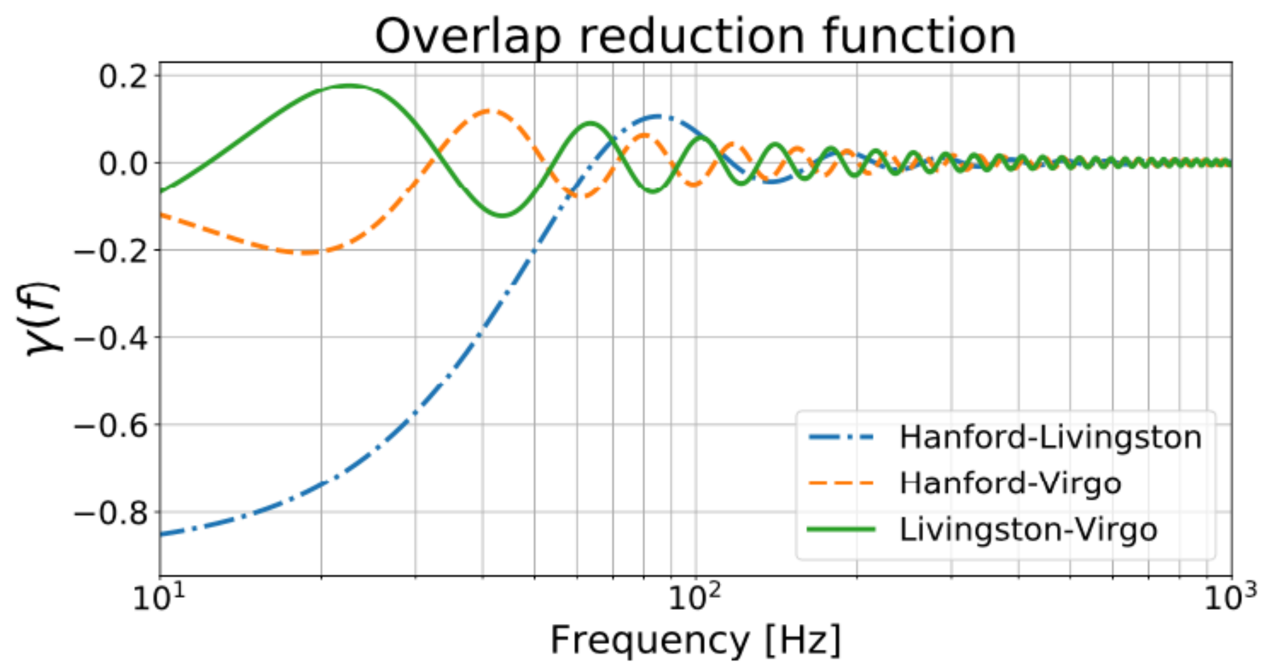
which implies...

$$\boxed{\left\langle \tilde{h}_H^*(f)\tilde{h}_L(f') \right\rangle = \frac{3H_0^2}{20\pi^2}\frac{\Omega_{GW}(f)}{f^3}\gamma(f)\delta(f-f')}$$

This means that we have direct access to $\Omega_{GW}(f)$!

Calculating the SGWB SNR

The detector geometry and antenna pattern is encoded in the *overlap reduction function*



$$\gamma(f) \equiv \frac{5}{8\pi} \sum_A \int d\hat{\Omega}\, e^{2\pi i f\, \hat{\Omega}\cdot(\mathbf{x}_H - \mathbf{x}_L)/c} F_H^A(\hat{\Omega}) F_L^A(\hat{\Omega})$$

where $F(\hat{\Omega})$ is the projection of the GW vector onto the detector's response tensor. We can then find the SNR to be

$$SNR = \frac{\mu}{\sigma} = \frac{3H_0^2}{10\pi^2} \left[2T \int_0^\infty df\, \frac{\gamma^2(f)\Omega_{GW}^2(f)}{f^6 P_1(f) P_2(f)} \right]^{1/2}$$

Learning from the SNR

$$SNR = \frac{\mu}{\sigma} = \frac{3H_0^2}{10\pi^2} \left[2T \int_0^\infty df \, \frac{\gamma^2(f)\Omega_{GW}^2(f)}{f^6 P_1(f) P_2(f)} \right]^{1/2}$$

What this tells us about the isotropic SGWB search

1. If we observe for long enough, we will detect *something*

2. We are most sensitive at low frequencies

3. SNR depends on model we pick. How to choose?

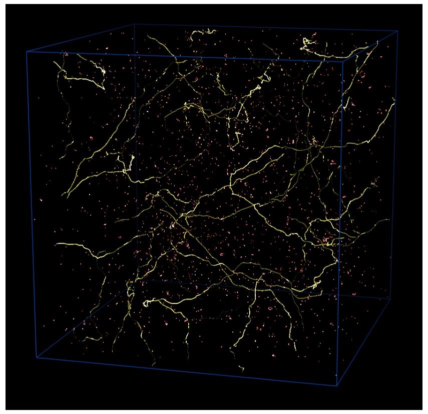
Choosing Models I

In general we assume a power law spectrum which allows us to write

$$\Omega_{GW}(f) = \Omega_0 \left(\frac{f}{f_{ref}} \right)^{\beta}$$

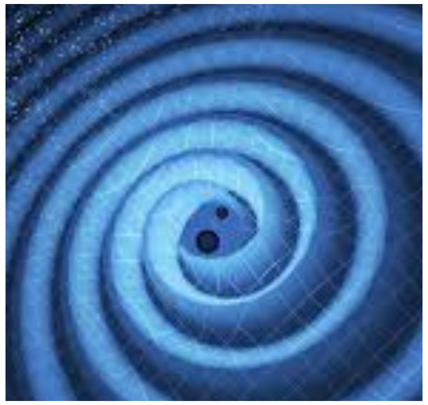
Currently Implemented Models

Cosmological Background



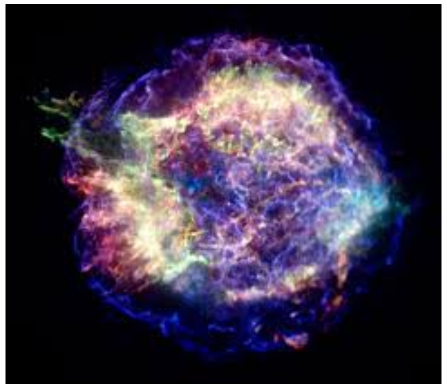
$\beta = 0$

Compact Binaries



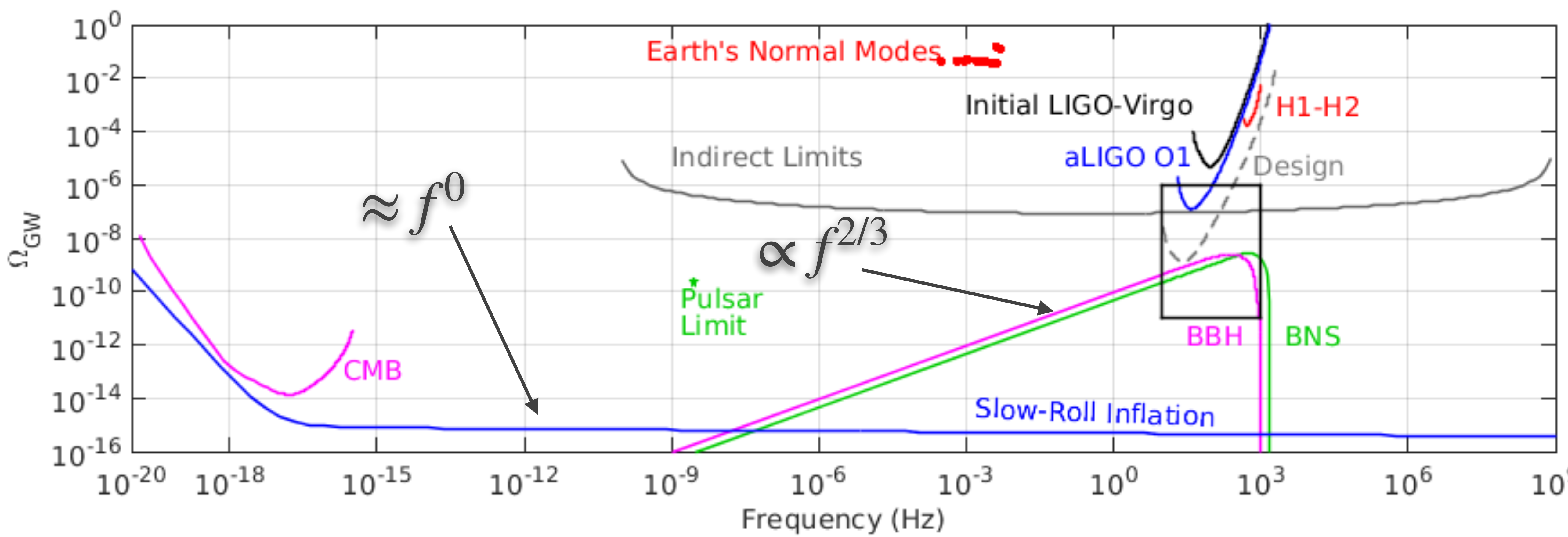
$\beta = 2/3$

Supernova



$\beta = 3$

Choosing Models II

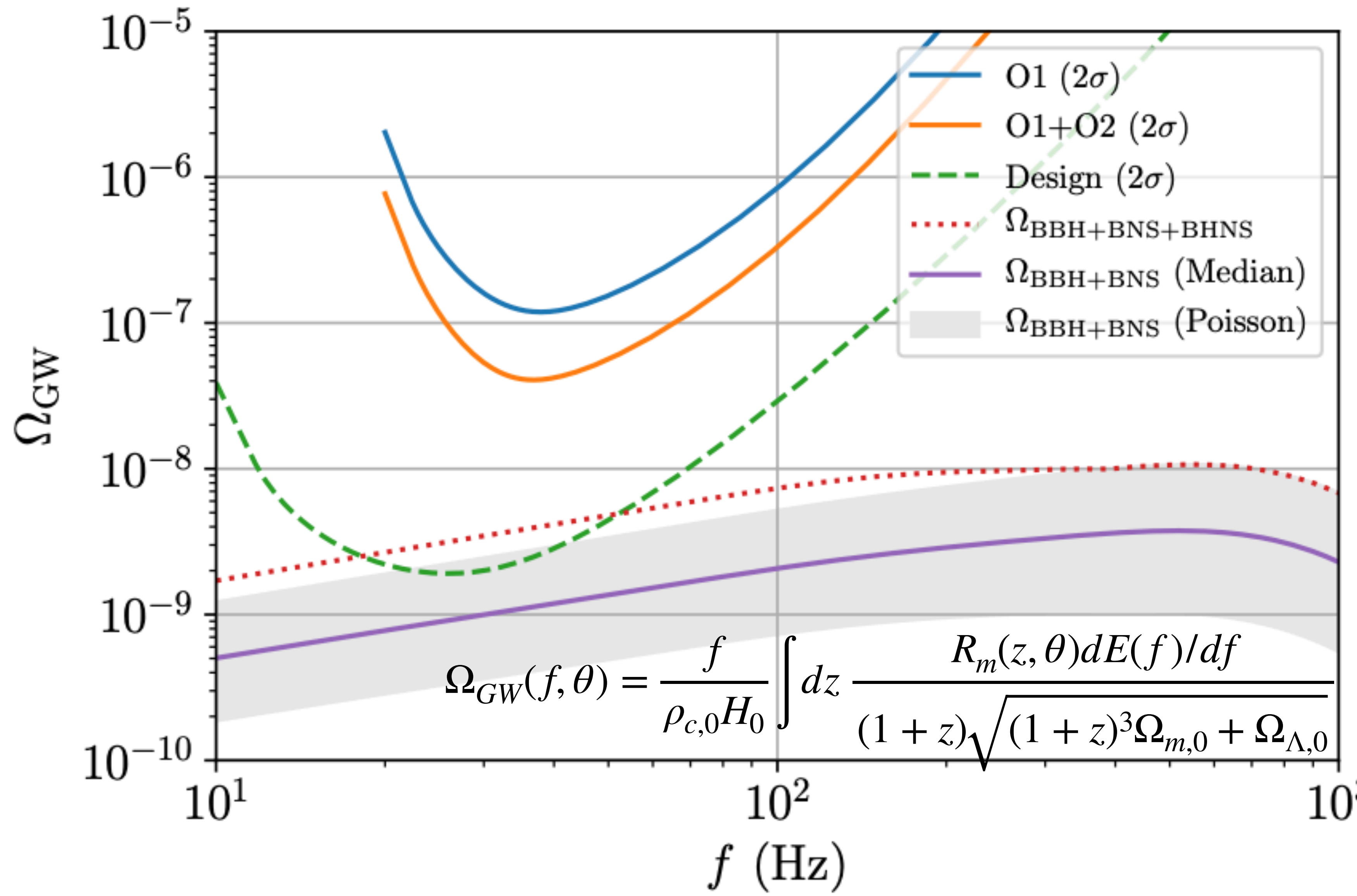


PI Integrated curves for various models of the stochastic GW background as well as LIGO & Virgo sensitivities. The indirect limits refer to constraints from CMB temperature and polarization power spectra, BAOs and BBN.

Ref: [Lasky et. al]

CBC Model

Sensitivity curves for O1, combined O1+O2, and design sensitivity. A power law stochastic background which lies tangent to one of these curves is detectable with 2σ significance. Design sensitivity assumes that the LIGO noise curve is determined by fundamental noise sources only.



Ref: Phinney et. al 2001

Probing the SGWB I

Collect H1 / L1 Timeseries

Raw (or Wiener Filtered) strain data

Timeseries Data Quality Cuts

Downsample to 4096Hz
Highpass 16th Order Butterworth filter (11 Hz knee frequency)
Reject times with "Category 1" vetos i.e., times with known glitches
Removes about 20% of data

Frequency Data Quality Cuts

Take FFTs of 192s segments and coarse grain down to 1/32 Hz
Reject "lines" if possible
Perform "delta-sigma" cut
Removes about 15% of data

Clean Data

Cross correlate data streams from Hanford and Livingston

Probing the SGWB II

Define cross-correlation statistic and variance for each frequency bin

$$\hat{C}(f) = \frac{20\pi^2 f^3}{3H_0^2 T} \frac{\mathrm{Re}[\tilde{s}_L^*(f)\tilde{s}_H(f)]}{\gamma(f)}$$

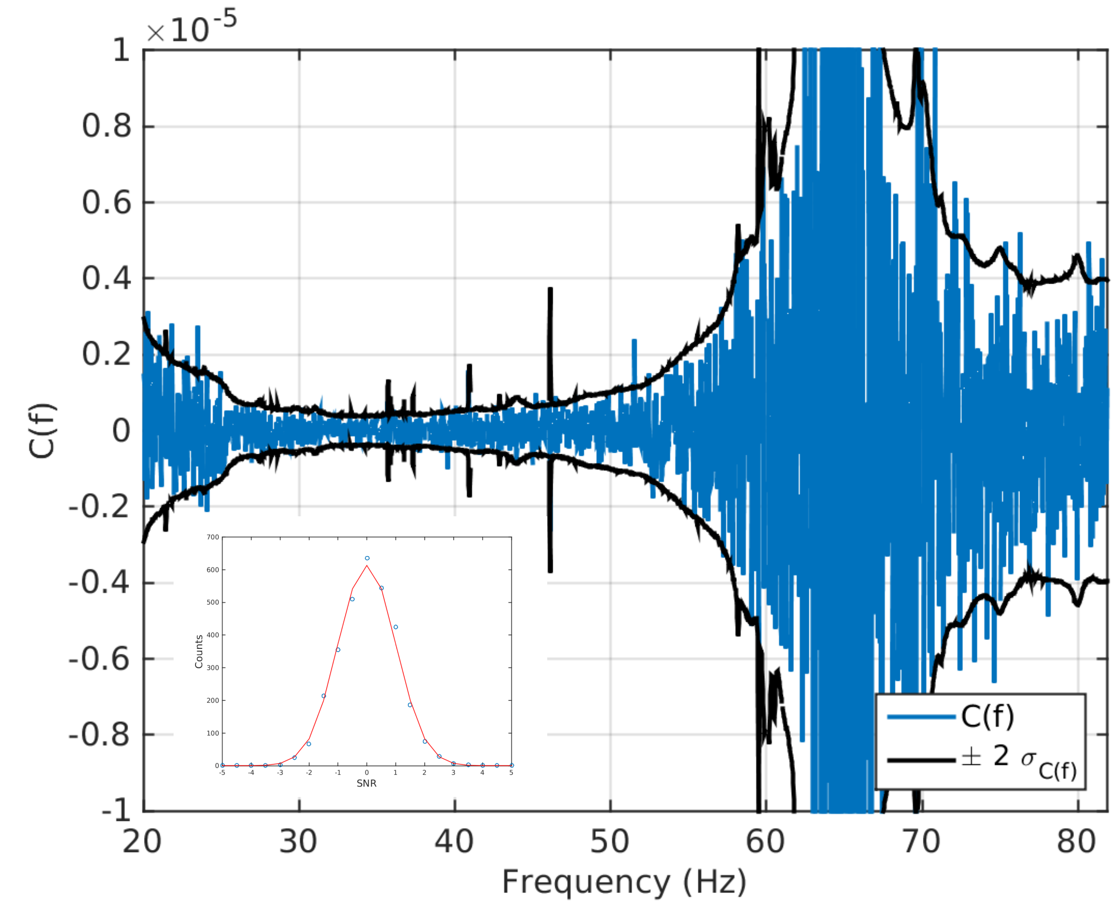
$$\sigma^2(f) = \left(\frac{10\pi^2}{3H_0^2}\right)^2 \frac{f^6}{2T\Delta f} \frac{P_1(f)P_2(f)}{\gamma^2(f)}$$

This gives the *optimal detection statistic*

$$\hat{\Omega}_{ref} = \frac{\sum_k w(f_k)^{-1}\hat{C}(f_k)\sigma^{-2}(f_k)}{\sum_k w(f_k)^{-2}\sigma^{-2}(f_k)}$$

where

$$\sigma_\Omega^{-2} = \sum_k w(f_k)^{-2}\sigma^{-2}(f_k) \quad \text{and} \quad w(f) = \frac{\Omega_{GW}(f_{ref})}{\Omega_{GW}(f)}$$



Cross-correlation statistic for LIGO during the second observing run. Loss of sensitivity at 64Hz is from a zero in the overlap reduction function

Isotropic SGWB Search Results

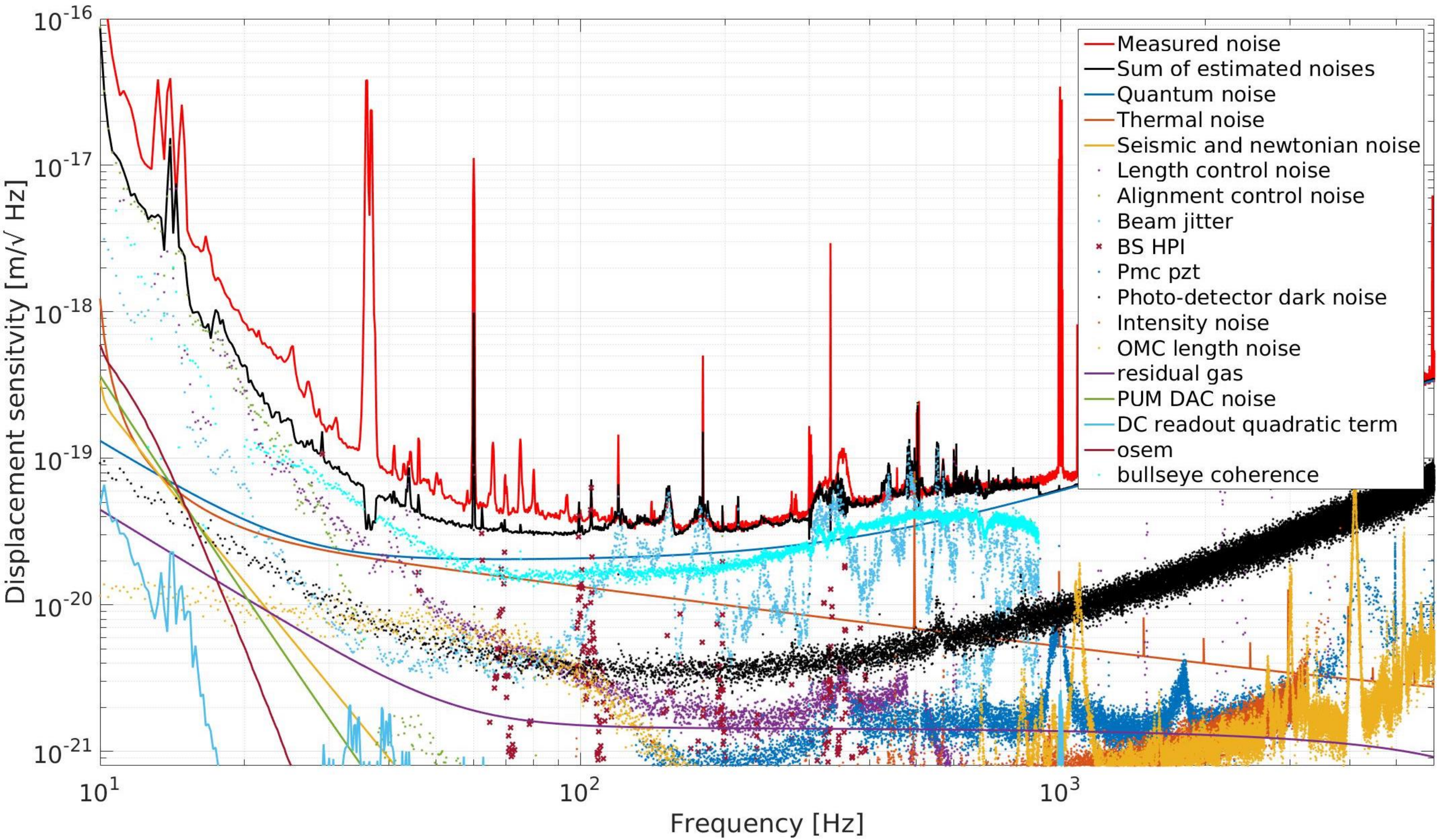
α	$\hat{\Omega}_{\rm ref}$ (O2)	$\hat{\Omega}_{\rm ref}$ (O1)	O2 Sensitive band
0	$(2.2 \pm 2.2) \times 10^{-8}$	$(4.4 \pm 6.0) \times 10^{-8}$	20-85.8 Hz
2/3	$(2.0 \pm 1.6) \times 10^{-8}$	$(3.5 \pm 4.4) \times 10^{-8}$	20-95.2 Hz
3	$(3.5 \pm 2.8) \times 10^{-9}$	$(3.7 \pm 6.6) \times 10^{-9}$	20-305 Hz

Point estimate with 1σ uncertainties for the optimal estimator for O1 and O2. The "sensitive band" contains 99% of the sensitivity for the shown spectral index.

Still ~1 order of magnitude above the CBC PI curve at ~30 Hz

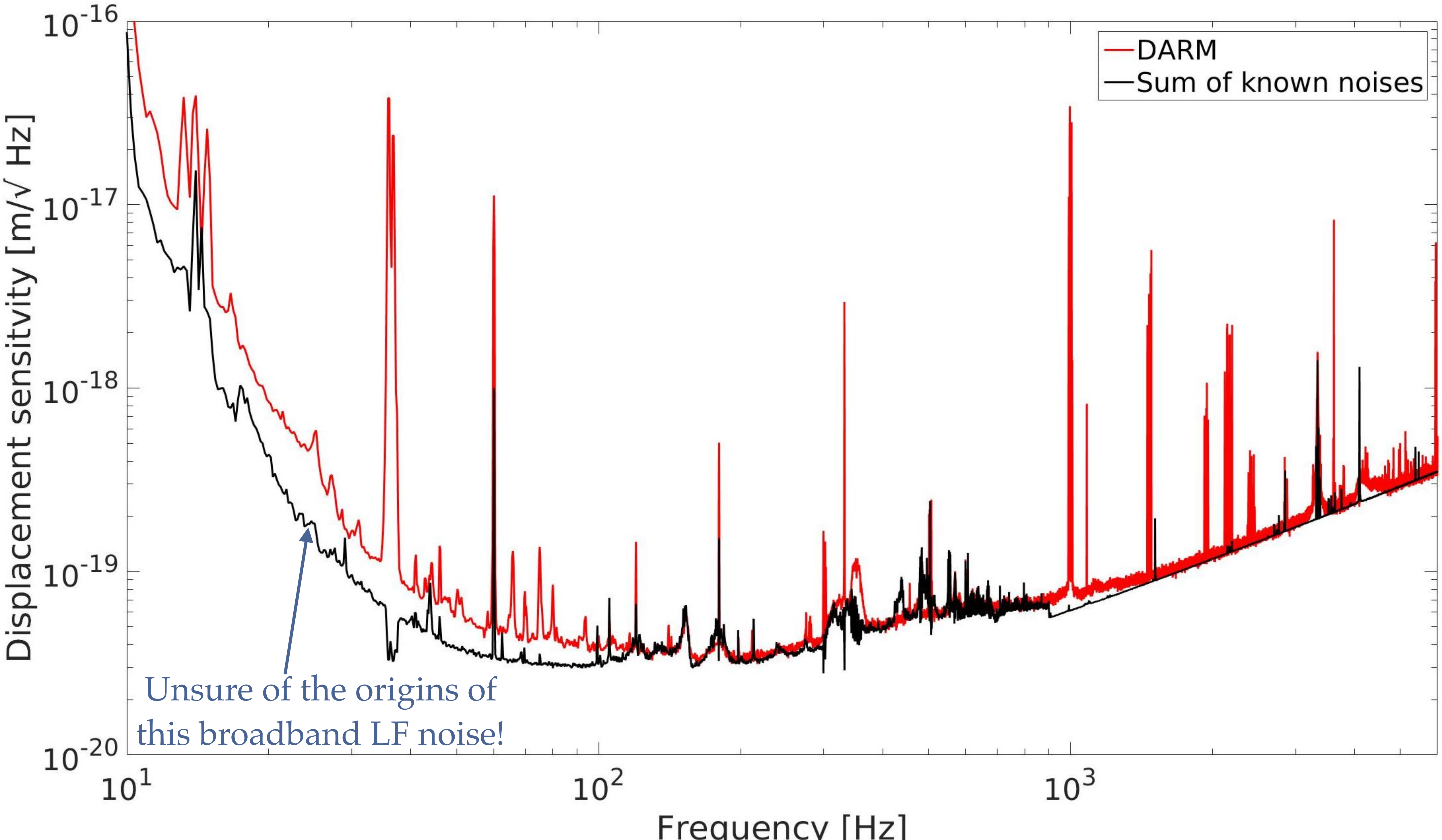
Enhancing Sensitivity Via Machine Learning

LIGO O2 Noise Budget



Ref: [Ronaldas Macas]

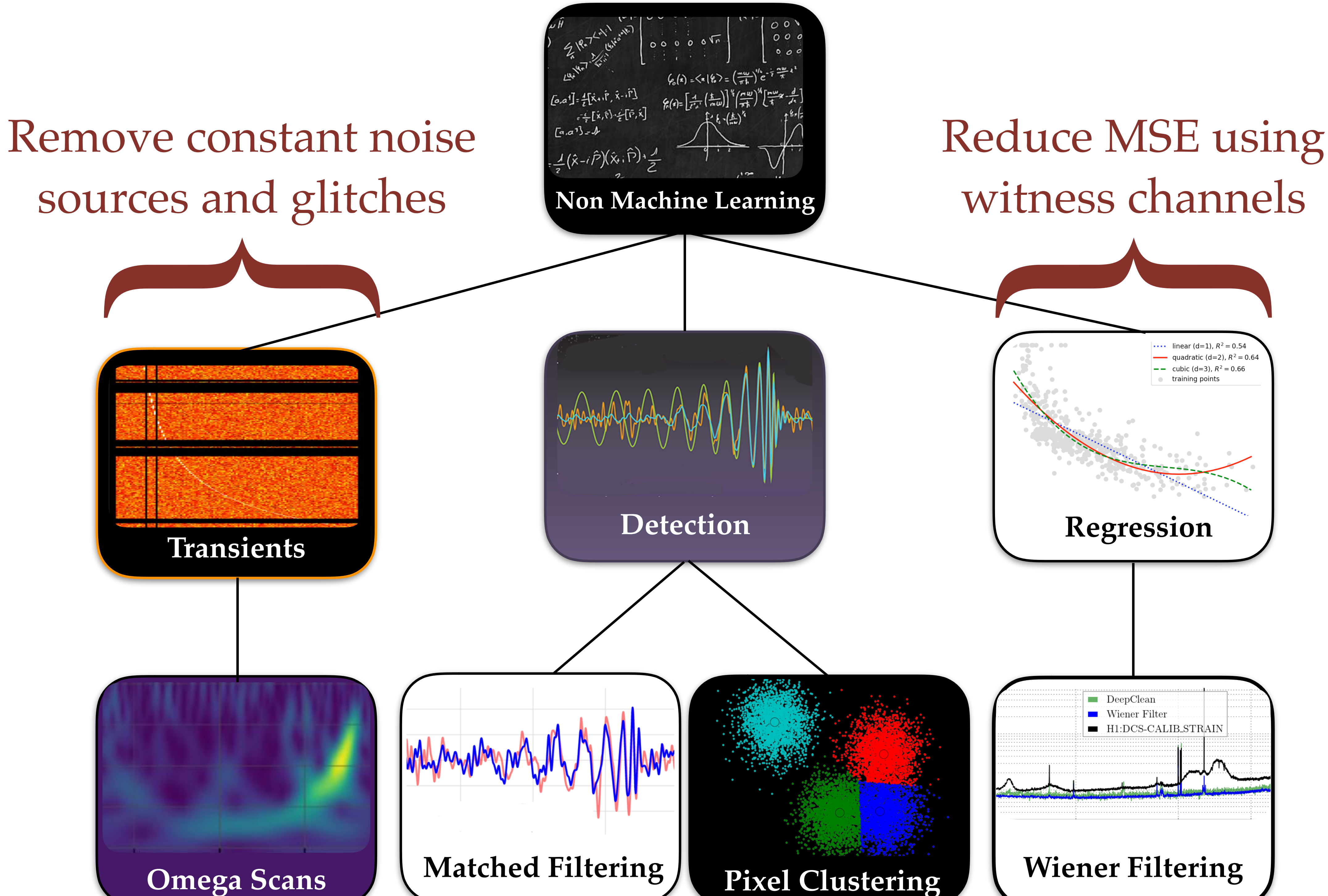
O2 Noise Budget - A Closer Look



Ref: [Detecting a stochastic background of gravitational radiation]

Ref: [Ronaldas Macas]

Data Quality



MISO - Wiener Filtering

$$\underbrace{\boxed{e[t]}}_{\text{"Error"}} = \underbrace{\boxed{s[t]}}_{\text{DARM}} - \underbrace{\boxed{\sum_i a_i \underbrace{\boxed{w[t-i]}}_{\text{Witnesses}}}}_{\text{Noise}}$$

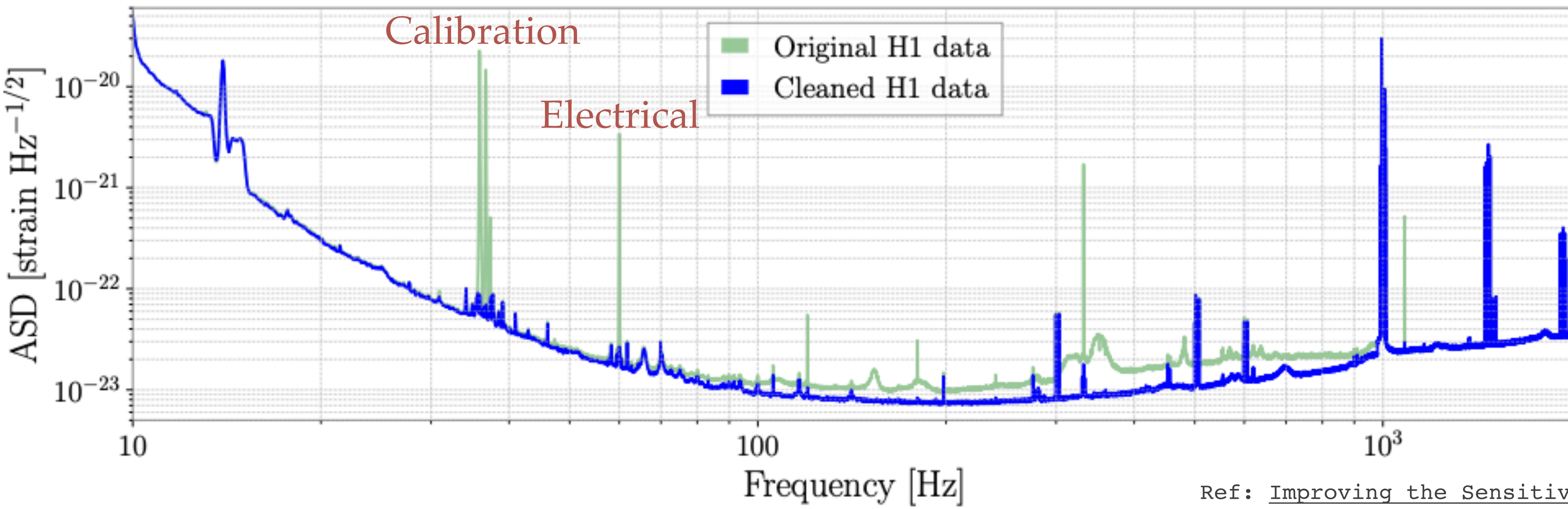
Estimate the transfer function by minimizing the MSE:

$$\frac{\partial}{\partial a_j}\left\langle e[t]^2 \right\rangle = 0$$

$$= 2\sum_j \langle w[t-j]w[t-i]\rangle\, a_j - 2\,\langle w[t-i]s[t]\rangle$$

Eq. Ref: MIT Press

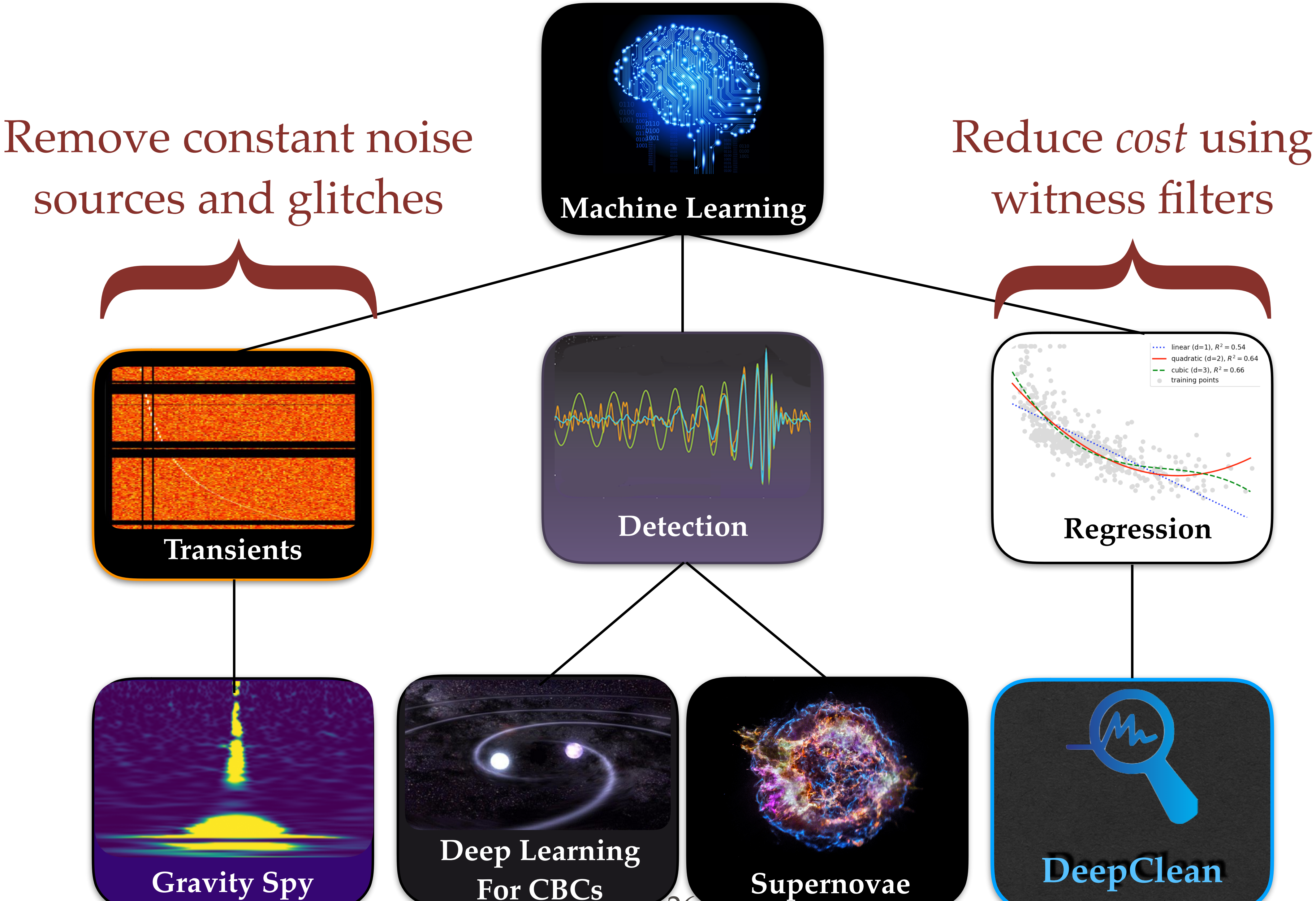
Wiener Filtering Results



Ref: Improving the Sensitivity of Advanced LIGO Using Noise Subtraction

- Filtering was run from 10 - 1024 Hz
- Essentially no subtraction below 80 Hz
- Subtracted noise from ~100 - 1000 Hz is primarily beam jitter

Machine Learning



How MLAs Work - I

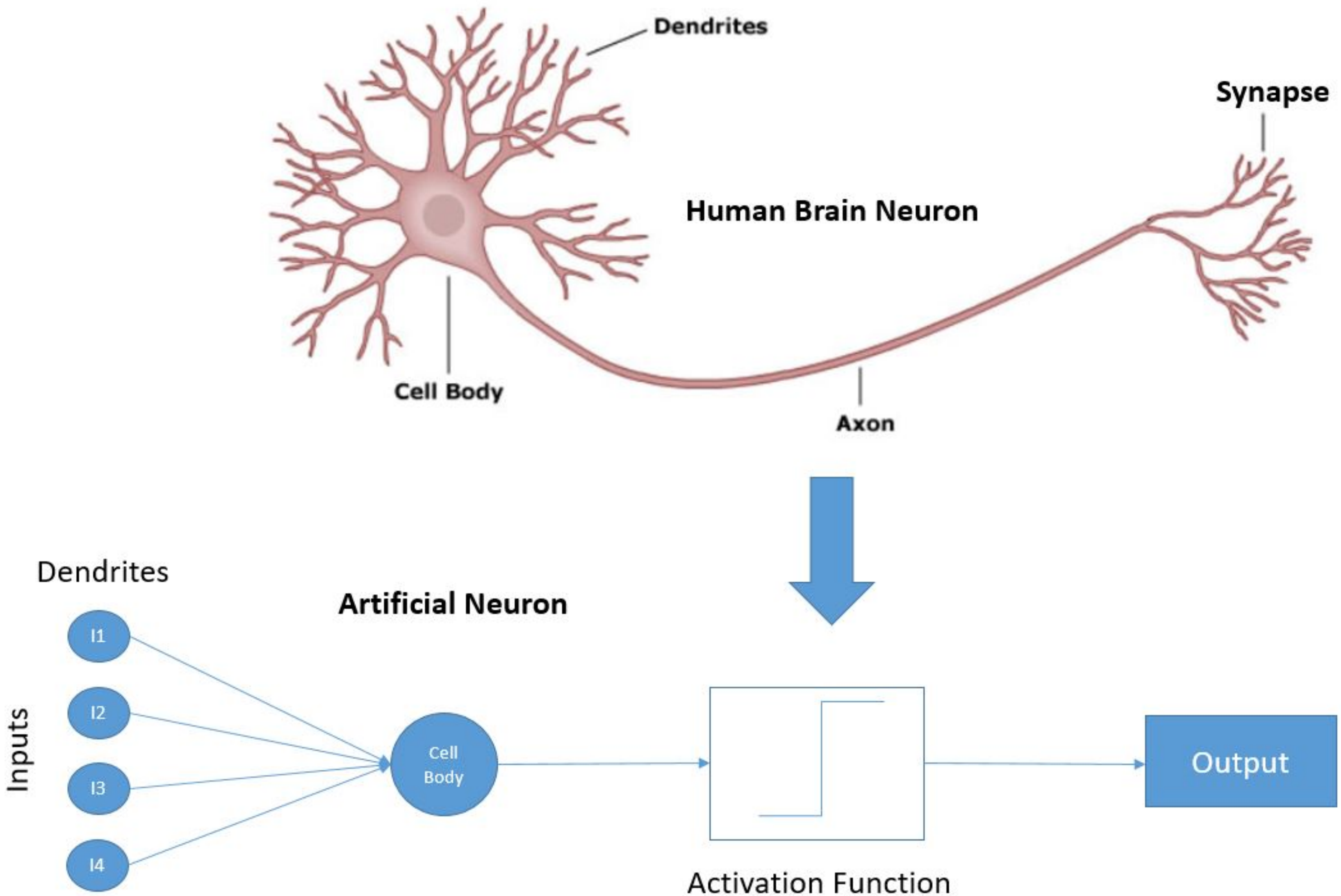
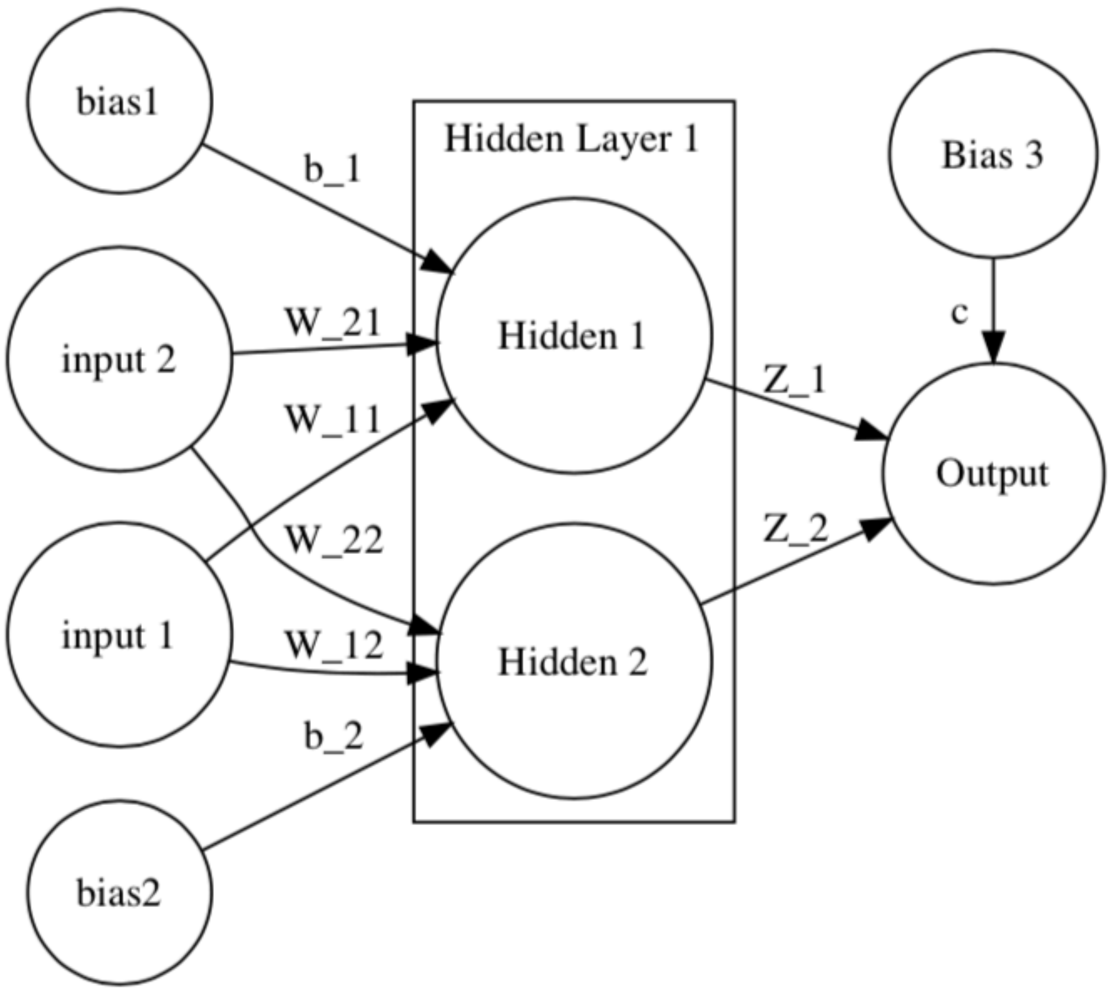


Image Credit: MPLSVPN - Moving Towards SDN and NFV Based Networks

How MLAs Work - 2



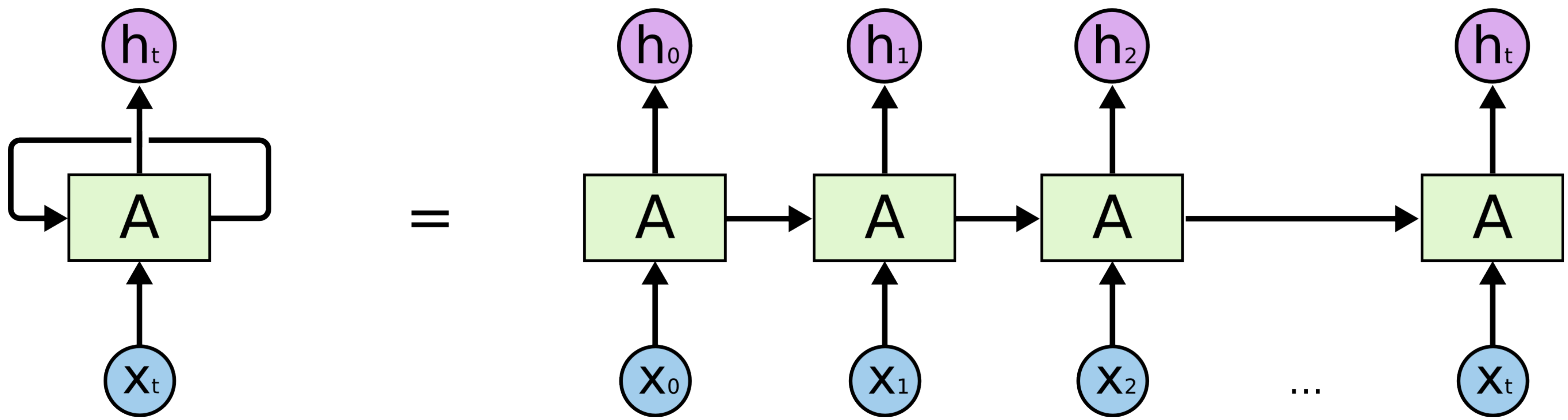
Activation Functions Input Data Biases

$$= g\left(Z^T f(W^T \mathbf{i} + \mathbf{b}) + c\right)$$

Weights

- Vanilla feed-forward network. The network is comprised of the hidden layer, activation functions, weights, and biases.
- **We** construct the architecture of the hidden layer and activation functions.
- **The network's** goal is to learn the values of the weights and biases.

Recurrent Neural Networks I

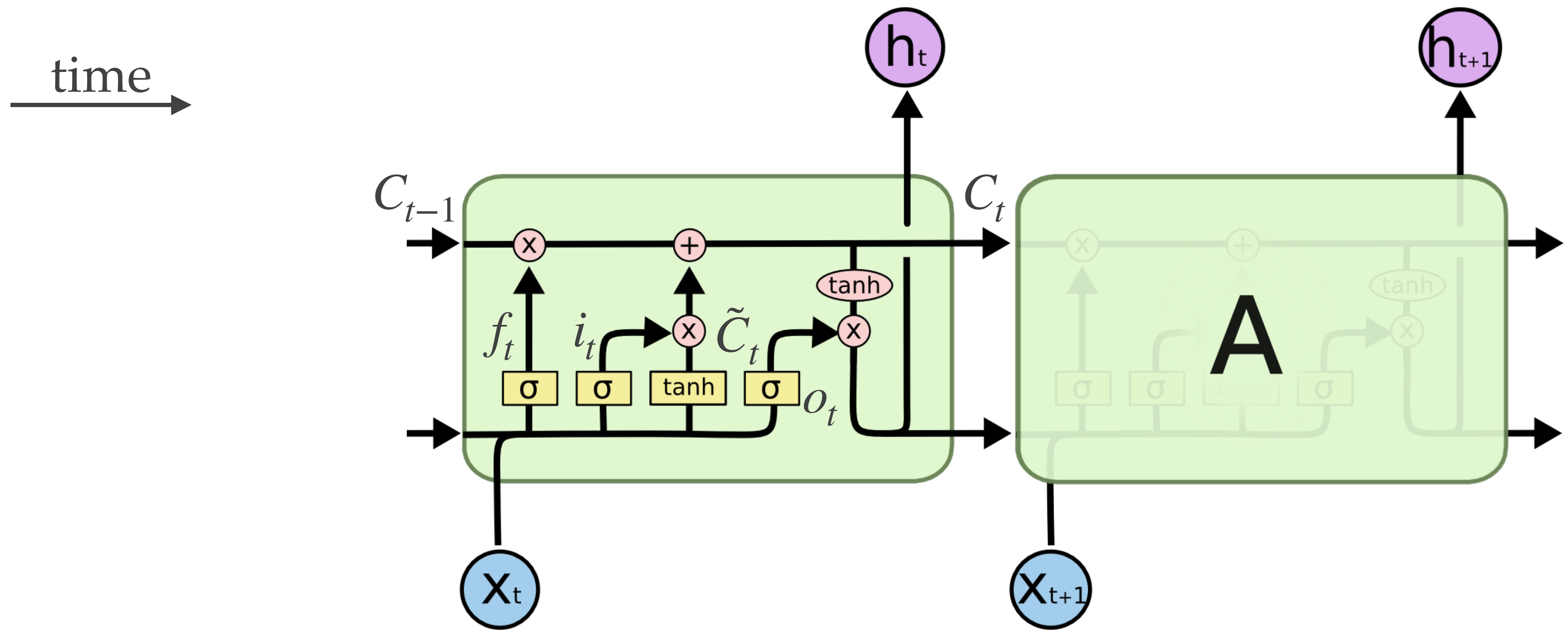


Each "A" block is a feed-forward network which receives previous output and current data as input. Therefore, past and present data together informs the predictions. In this sense, the network has memory.

Expanding on this leads to Long Short-Term Memory (LSTM) Networks.

Image credit: Chris Olah

Recurrent Neural Networks II

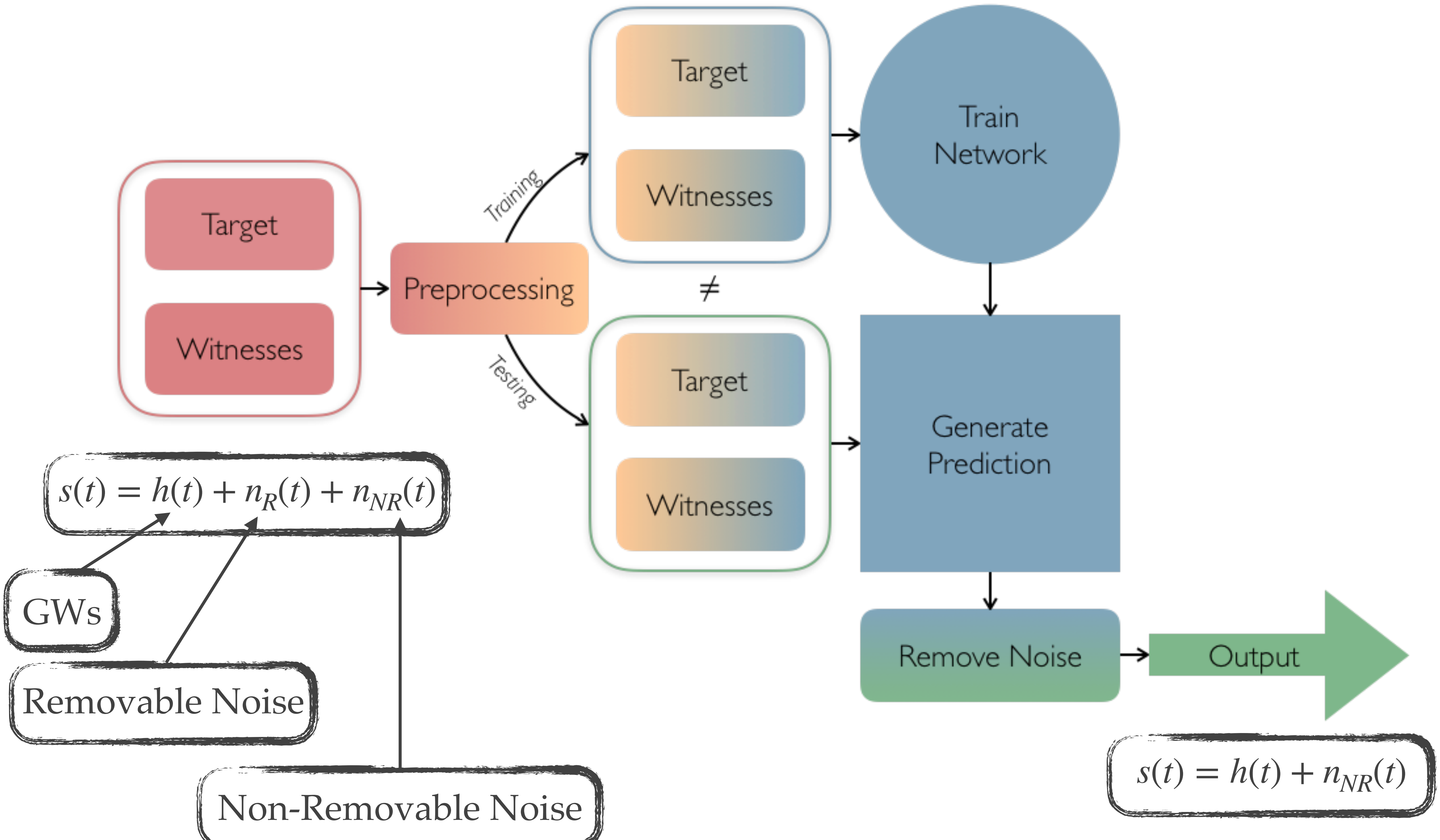


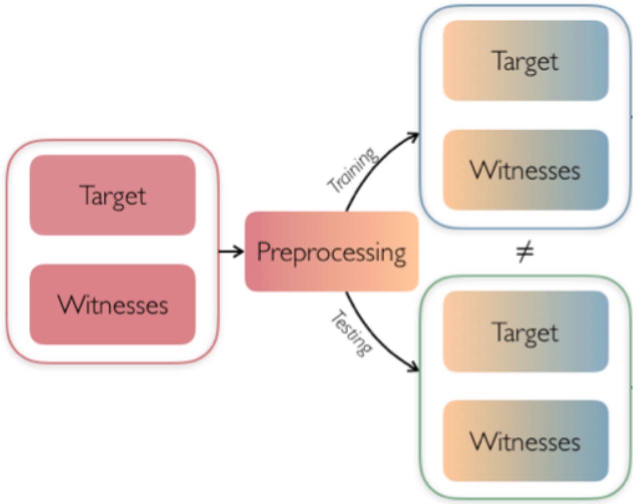
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot \tanh(C_t)$$

Each input time step is analyzed through a series of gates (forget, input, output) and this information is fed into a "cell state."

Image credit: [Chris Olah]()

DEEP CLEAN Workflow I





Workflow II

- Separate out the target from the reference data.

- Preprocessing can involve many steps. Generally, we normalize (or standardize):

$$Data \rightarrow \frac{Data - \mu}{\sigma}$$

To reduce training time and increase performance, we also bandpass the data into several bands and "loop" over the network performing subtraction in each band separately

- For supervised learning, split the data into training and testing samples. Network never "sees" testing data

Workflow III

Network training consists (roughly) of three parts:

1. Given the witnesses $(\bar{w}_t, \ldots \bar{w}_{t-L})$ and lookback L, calculate a prediction $\tilde{n}_{R_t}$ (repeat for each time step)

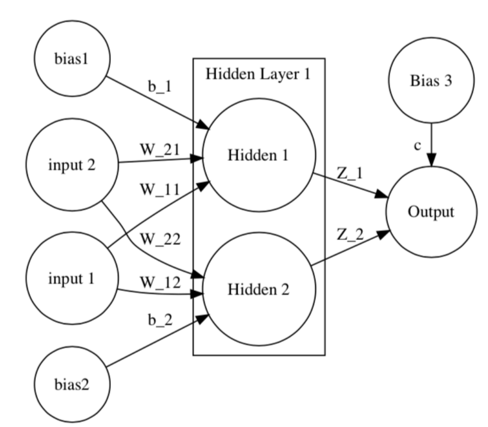
$$Network[\bar{w}_t, \bar{w}_{t-1}, \ldots \bar{w}_{t-L}] = \tilde{n}_{R_t} \qquad \big(s(t) = h(t) + n_R(t) + n_{NR}(t)\big)$$

2. Calculate the error of the prediction with the target value through a cost function $J(s_t, \tilde{n}_{R_t})$. Typically use MSE

$$J = \mathcal{N} \sum_t J(s_t, \tilde{n}_{R_t}) = \mathcal{N} \sum_t \left(s_t - \tilde{n}_{R_t}\right)^2$$

3. Update the weights to minimize the cost function

$$\bar{Z} \rightarrow \bar{Z} - \eta \bar{\nabla} J(s_t, \tilde{n}_{R_t})$$



Training Without a Formal Target I

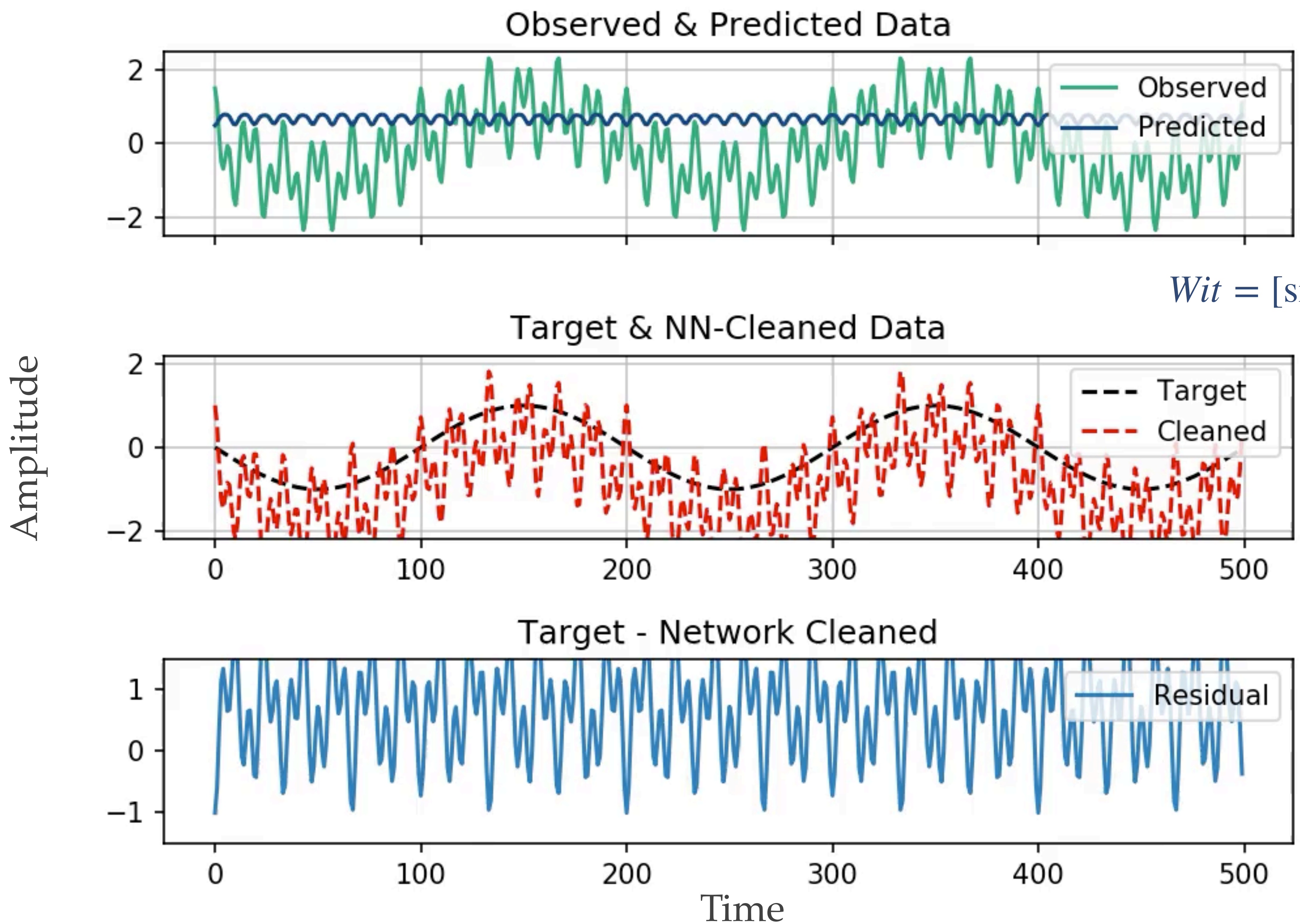
Question:
How does the network "know" to only subtract noise and not to remove any of the true signal?

Answer:
The witness channels will not have GWs imprinted upon them (no way to measure GWs with a seismometer & magnetometer)

The neural network learns on (nearly) stationary noise sources; short transients such as glitches and GWs will not be learned

However, continuous wave sources which match the frequency of a witness channel will be removed!

Training Without a Formal Target II



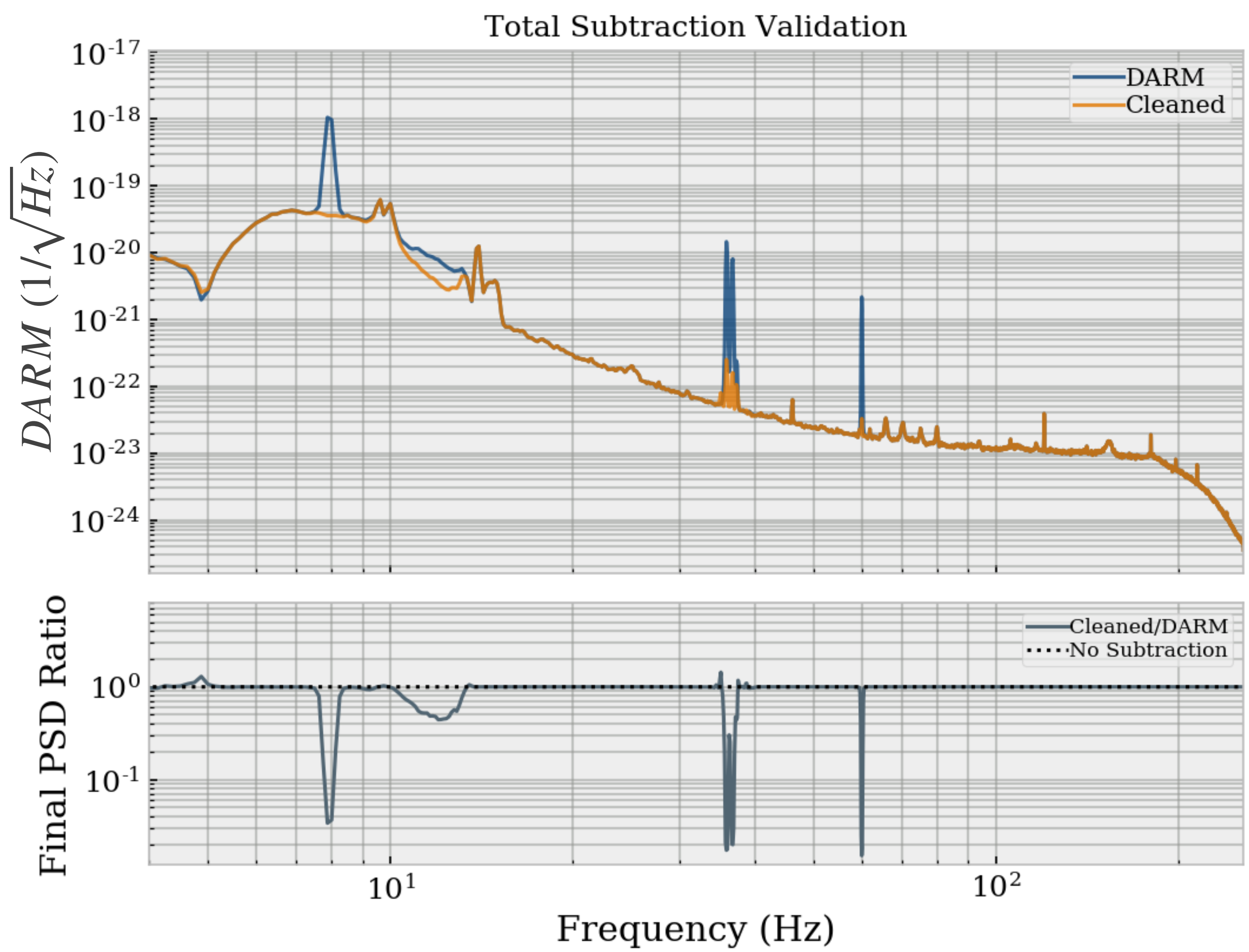
$Obs = \sin(a) + \sin(b)\sin(c)$

$Wit = [\sin(b), \sin(c)]$

LHO O2 Calibration Lines

O2 Channel List

GDS-CALIB_STRAIN
PSL-DIAG_BULLSEYE_PIT_OUT_DQ
PSL-DIAG_BULLSEYE_YAW_OUT_DQ
PSL-DIAG_BULLSEYE_WID_OUT_DQ
MC-WFS_A_DC_PIT_OUT_DQ
IMC-WFS_B_DC_PIT_OUT_DQ
IMC-WFS_A_DC_YAW_OUT_DQ
IMC-WFS_B_DC_YAW_OUT_DQ
ASC-DHARD_P_OUT_DQ
ASC-DHARD_Y_OUT_DQ
ASC-CHARD_P_OUT_DQ
ASC-CHARD_Y_OUT_DQ
LSC-CAL_LINE_SUM_DQ
LSC-SRCL_IN1_DQ
LSC-MICH_IN1_DQ
LSC-PRCL_IN1_DQ
PEM-EY_MAINSMON_EBAY_1_DQ
PEM-EY_MAINSMON_EBAY_2_DQ
PEM-EY_MAINSMON_EBAY_3_DQ
CAL-CS_LINE_SUM_DQ
CAL-PCALY_TX_PD_OUT_DQ
CAL-PCALY_EXC_SUM_DQ
SUS-ETMY_L3_CAL_LINE_OUT_DQ

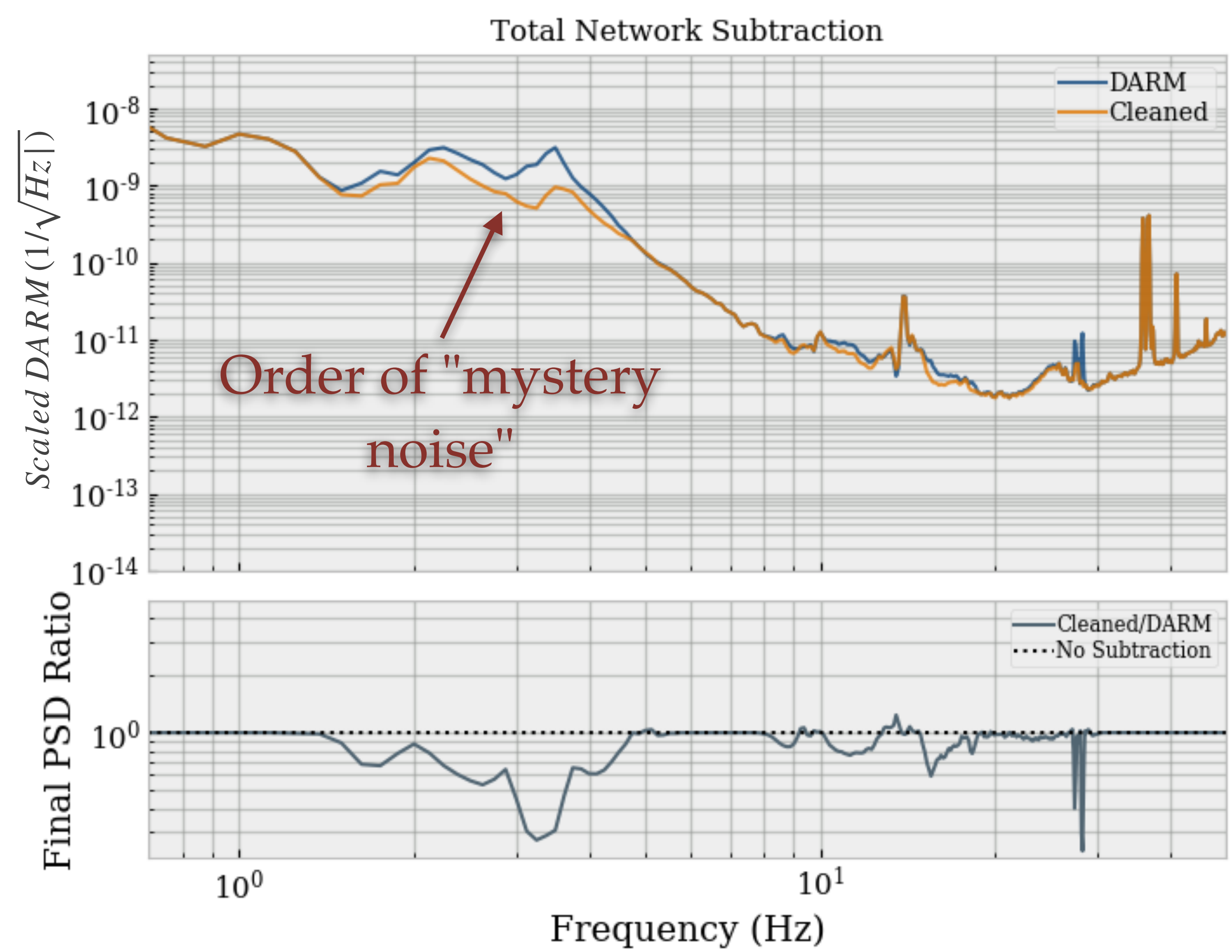


Removing calibration lines and 60 Hz mains from LHO during O2 (+ more?)

LHO O1 GW150914

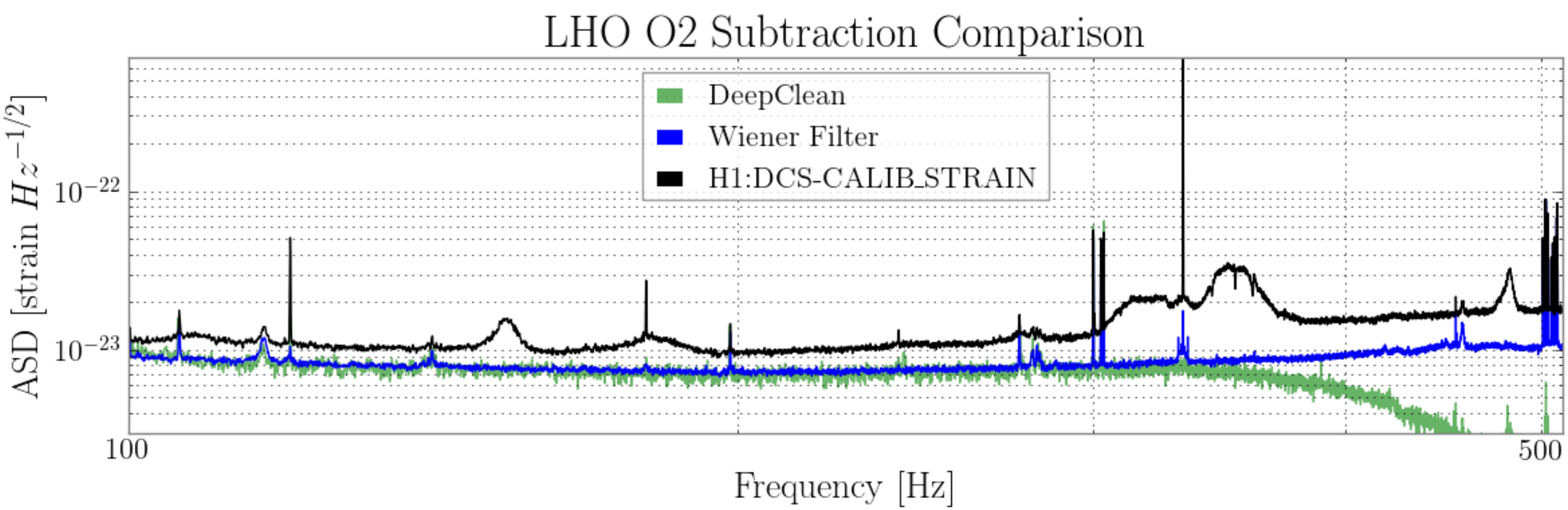
O1 Channel List

CAL-DELTAL_EXTERNAL_DQ
ASC-CHARD_P_OUT_DQ
ASC-CHARD_Y_OUT_DQ
ASC-DHARD_P_OUT_DQ
ASC-DHARD_Y_OUT_DQ
LSC-MICH_OUT_DQ
LSC-PRCL_OUT_DQ
LSC-SRCL_OUT_DQ
PEM-CS_ACC_HAM4_SR2_X_DQ
PEM-CS_ACC_HAM6_OMC_X_DQ
PEM-CS_ACC_HAM2_PRM_Y_DQ
PEM-CS_MIC_LVEA_INPUTOPTICS_DQ
PEM-CS_MIC_LVEA_OUTPUTOPTICS_DQ
ASC-X_TR_A_PIT_OUT_DQ
ASC-X_TR_A_YAW_OUT_DQ
ASC-X_TR_B_PIT_OUT_DQ
ASC-X_TR_B_YAW_OUT_DQ
ASC-Y_TR_A_PIT_OUT_DQ
ASC-Y_TR_A_YAW_OUT_DQ
ASC-Y_TR_B_PIT_OUT_DQ
ASC-Y_TR_B_YAW_OUT_DQ



Jitter subtraction at LHO using 1024s of data surrounding GW150914

Comparison to Wiener Filter

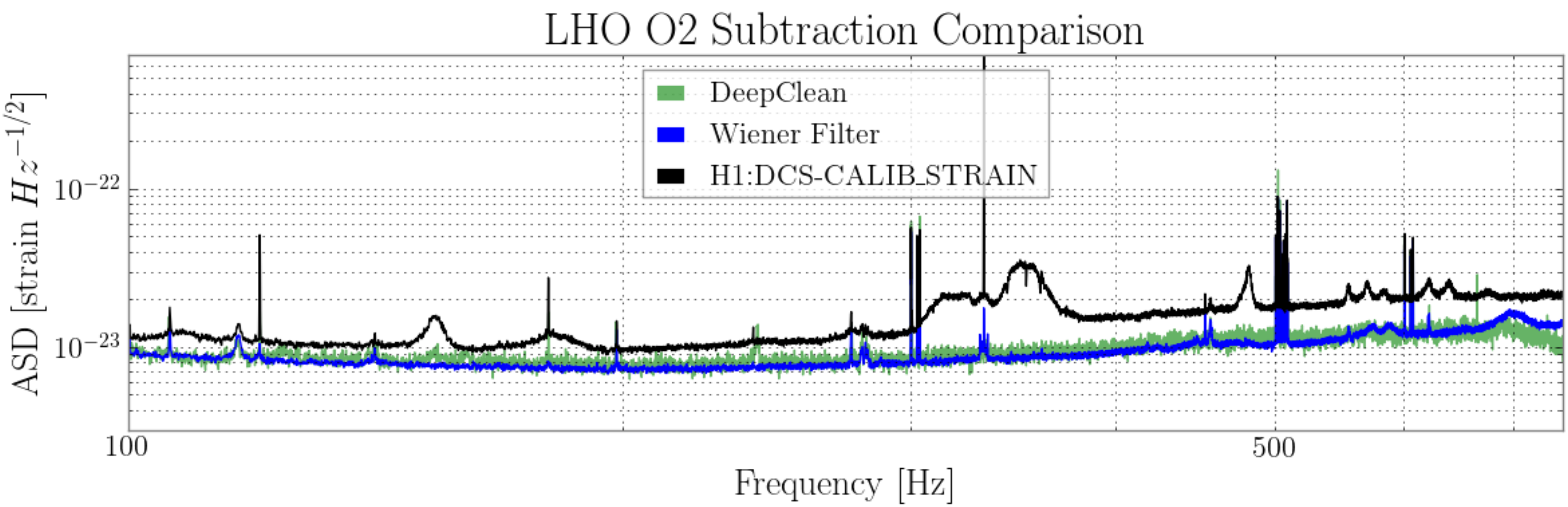


The broadband performance of DeepClean against the O2 linear subtraction is essentially identical (linear activation functions)

Validates the WF method and DeepClean's network.

Due to sample_rate = 1024 Hz, the amplitude falls off as we approach the Nyquist frequency.

Comparison to Wiener Filter



Same network as previous slide, but with sample_rate = 2048.

The network isn't fully converged yet (twice as much data would need to run a little longer) but the performance is still roughly the same.

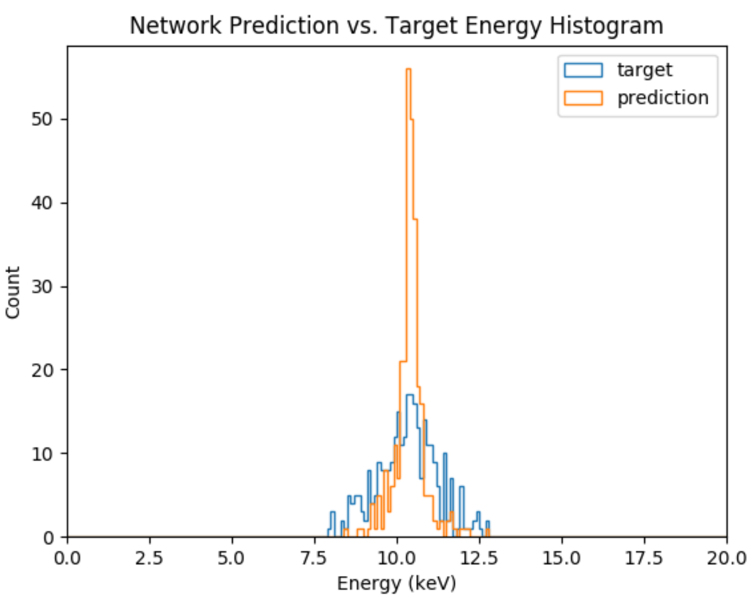
Next Steps for DeepClean

- Try to find the "right" channel list. Not easy!
- Determine where / why we beat the WF. Focus on 20-85 Hz to aid the stochastic search
- Do waveform injections followed by NN-cleaning and then perform parameter estimation. Compare to WF

THANK YOU!

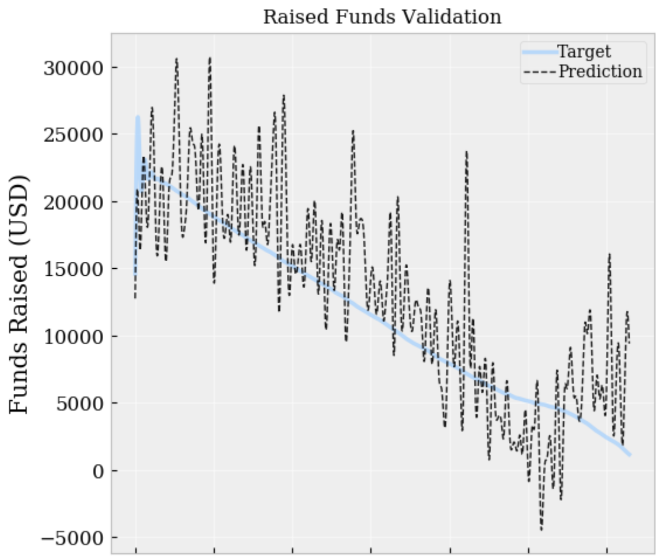
--- Extra Slides ---

Nonlinear MLAs Outside LIGO

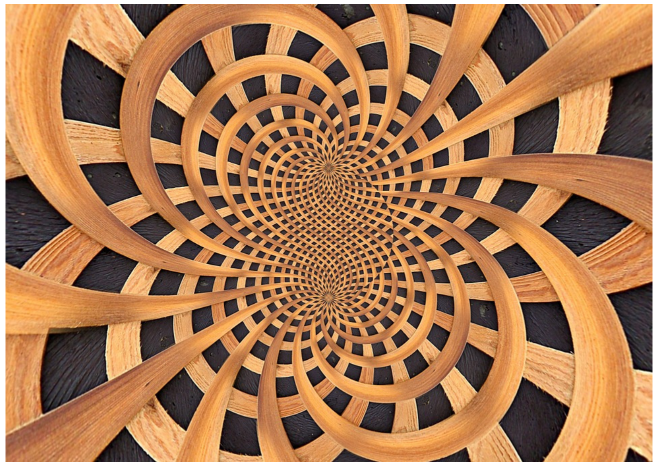


Reconstructing event energies in CDMS

Enticing jellyfish



Determining how much money a candidate needs to raise to win an election (yes, really)



+

=

Transverse Traceless Gauge

Under the transformation $x \to x' = x + \xi$, we find

$$h_{\mu\nu} \to h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \xi_\nu - \partial_\nu \xi_\mu$$

Using the trace reversed gauge where

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h$$

We find

$$\bar{h}_{\mu\nu} \to \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} + \partial_\mu \xi_\nu - \partial_\nu \xi_\mu - \eta_{\mu\nu}\eta^{\alpha\beta}\partial_\alpha \xi_\beta$$

We can choose the Lorenz gauge where

$$0 = \partial_{\mu'} \bar{h}^{\mu'}_{\nu} = \partial_\mu \bar{h}^\mu_\nu - \Box \xi_\nu$$

Lastly, we can pick $\Box \xi_\nu = 0$ which implies that

$$\partial_\mu h^{\mu\nu} = 0$$

From GR to GW

Solutions to the GW equation can be readily found in the near and far fields

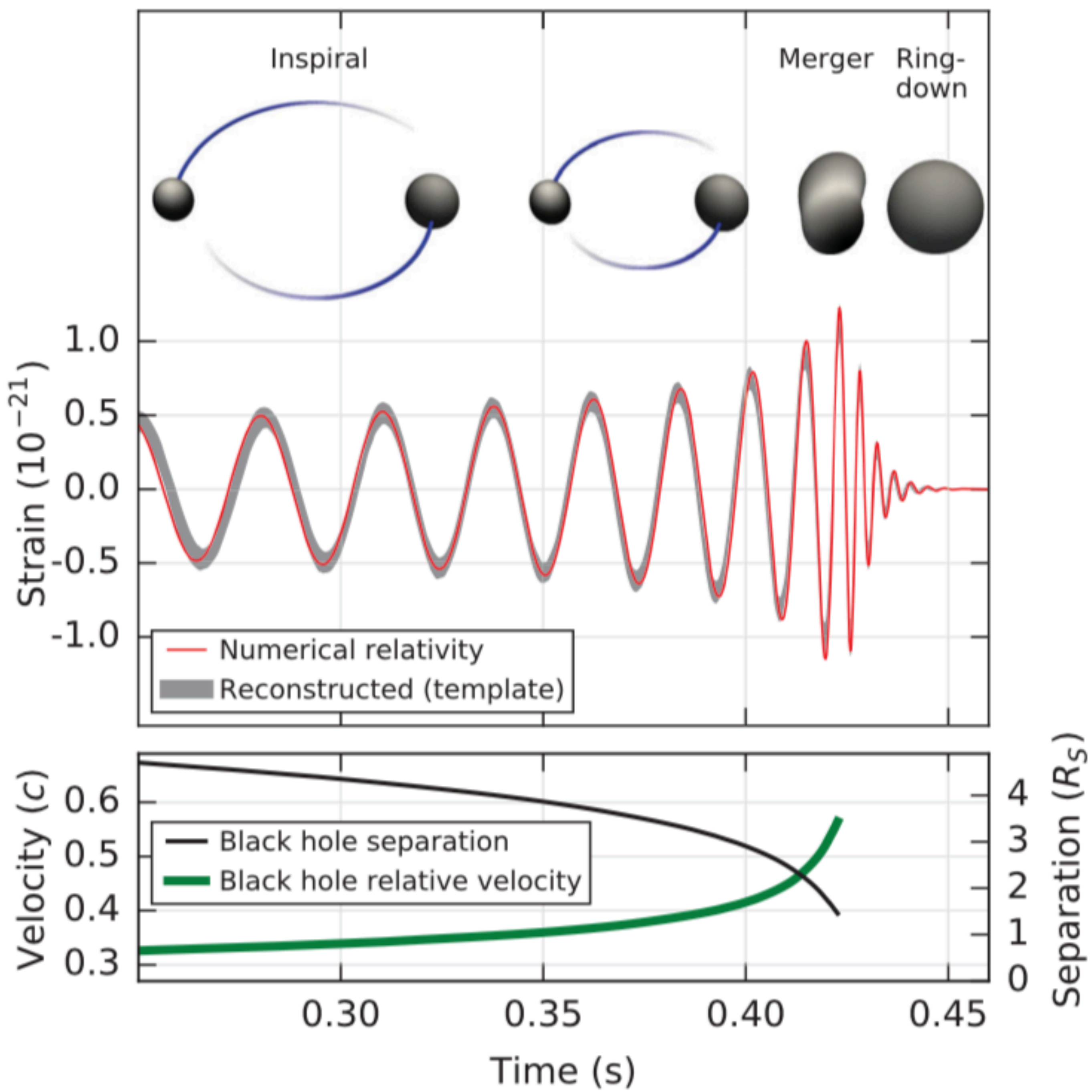
Far Field: $r \gg \lambda_{GW} \gg R_s$

$$\bar{h}_{ij}(t, \mathbf{x}) \simeq \frac{2G}{c^4 r} \ddot{I}_{ij}(t - r/c) \qquad I_{ij}(t) = \int x_i x_j T_{00}(t - r/c, \mathbf{x}) d^3\mathbf{x}$$

Near Field: $\lambda_{GW} \gg r \gg R_s$

$$\bar{h}_{ij}(t, \mathbf{x}) \simeq \frac{2G}{c^4 r} \ddot{\rlap{-}I}_{ij}(t - r/c) \qquad \rlap{-}I_{ij}(t) = \int (x_i x_j - \frac{1}{3}\delta_{ij} r^2) T_{00}(t - r/c, \mathbf{x}) d^3\mathbf{x}$$

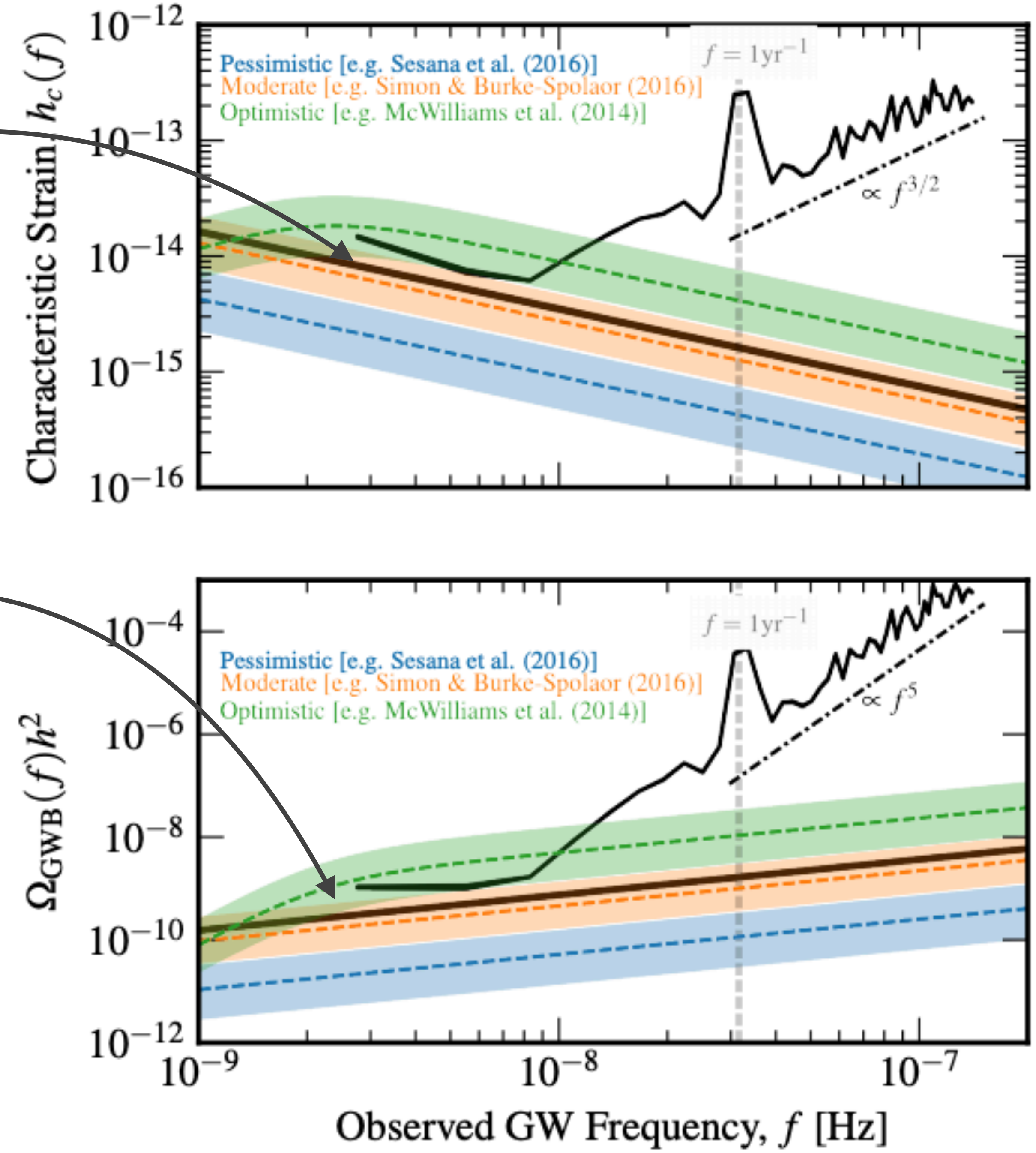
CBC GW Waveform



NanoGrav 11 Year Stochastic Paper

$$\Omega_{GW}(f) \propto f^{2/3} \propto f^3 S_h(f) \propto f^2 h_c^2(f)$$

$$\rightarrow h_c(f) \propto f^{-2/3}$$



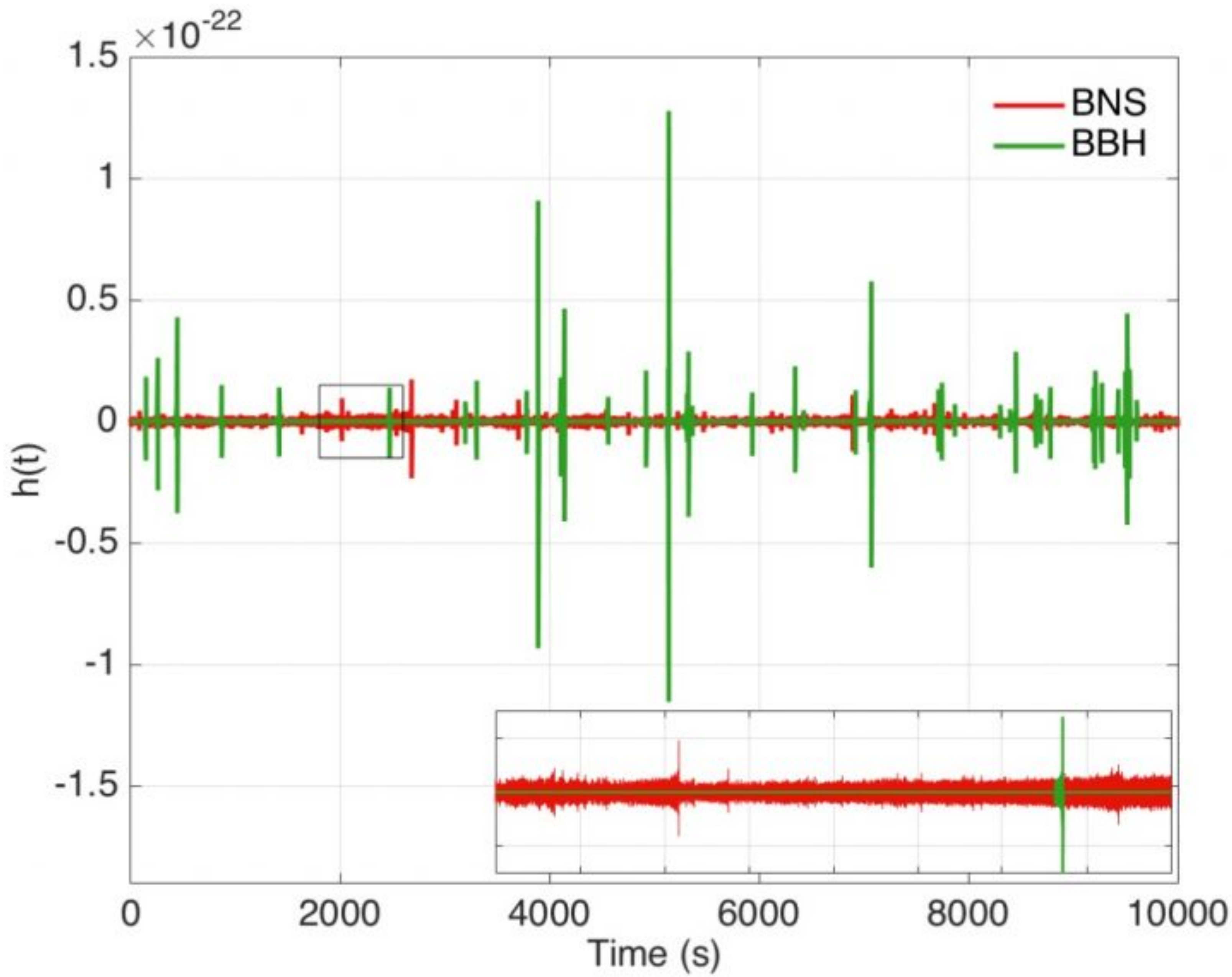
$$\Omega_{GW}(f) \propto f^{2/3}$$

Ref: arxiv 1801.02617

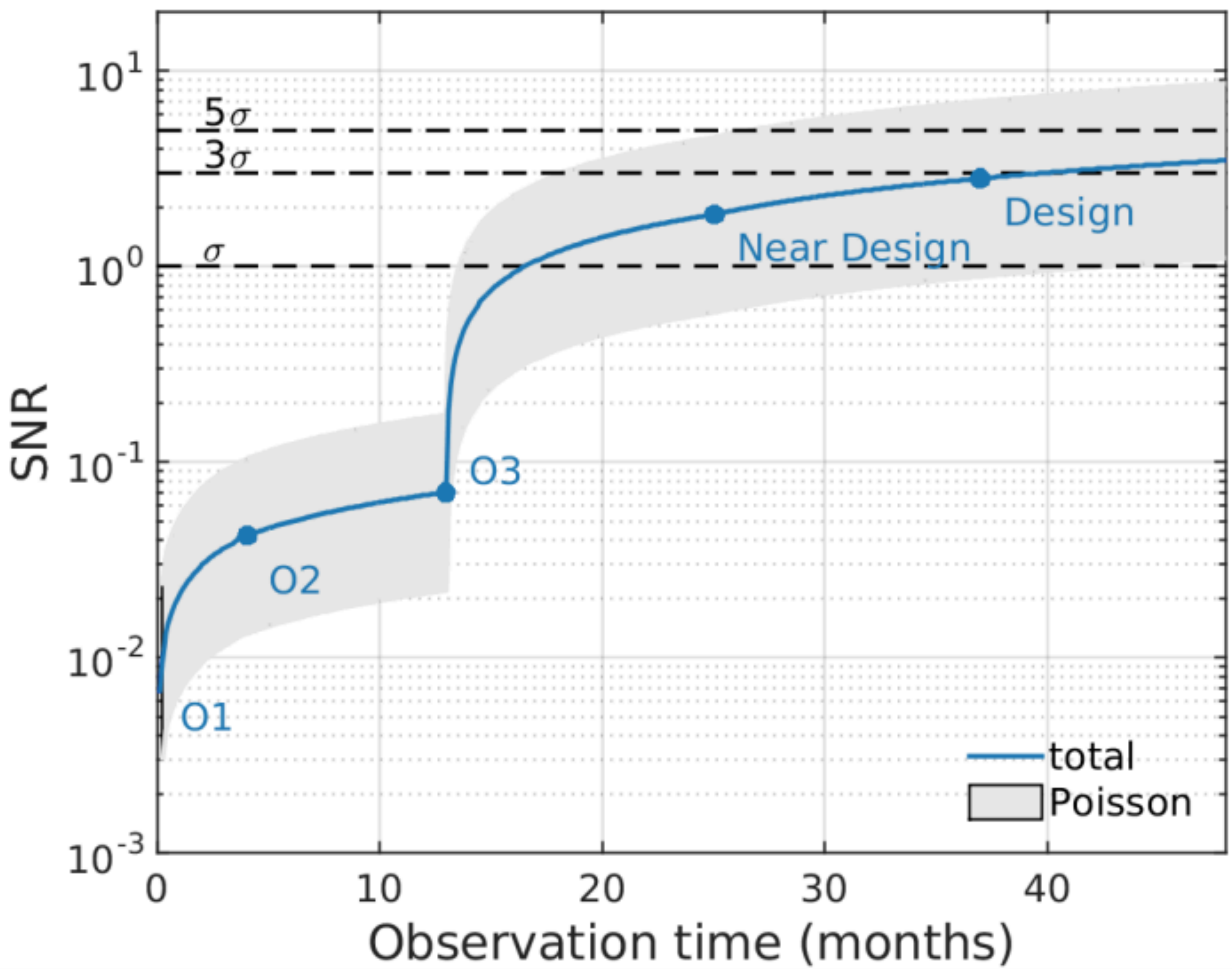
The Stochastic CBC Background

"Popcorn" BBH background and persistent BNS background.

[arxiv ref]



The Stochastic GW Background



Isotropic SGWB Bayesian Search

	Uniform prior		Log-uniform prior	
α	O1+O2	O1	O1+O2	O1
0	6.0×10^{-8}	1.7×10^{-7}	3.5×10^{-8}	6.4×10^{-8}
2/3	4.8×10^{-8}	1.3×10^{-7}	3.0×10^{-8}	5.1×10^{-8}
3	7.9×10^{-9}	1.7×10^{-8}	5.1×10^{-9}	6.7×10^{-9}
Marg.	1.1×10^{-7}	2.5×10^{-7}	3.4×10^{-8}	5.5×10^{-8}

TABLE II. 95% credible upper limits on $\Omega_{\rm ref}$ for different power law models (fixed α), as well as marginalizing over α, for combined O1 and O2 data (current limits) and for O1 data (previous limits) [66]. We show results for two priors, one which is uniform in $\Omega_{\rm ref}$, and one which is uniform in the logarithm of $\Omega_{\rm ref}$.

90% Confidence Upper Limits

$$0.90 = \int_{-\infty}^{\Omega_\alpha^{90}} p(\hat{\Omega}_\alpha|\Omega_\alpha)p(\Omega_\alpha)\, d\Omega_\alpha = \int_0^{\Omega_\alpha^{90}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\hat{\Omega}_\alpha - \Omega_\alpha)^2}{2\sigma_\alpha^2}}\, d\Omega_\alpha$$

Isotropic SGWB Models

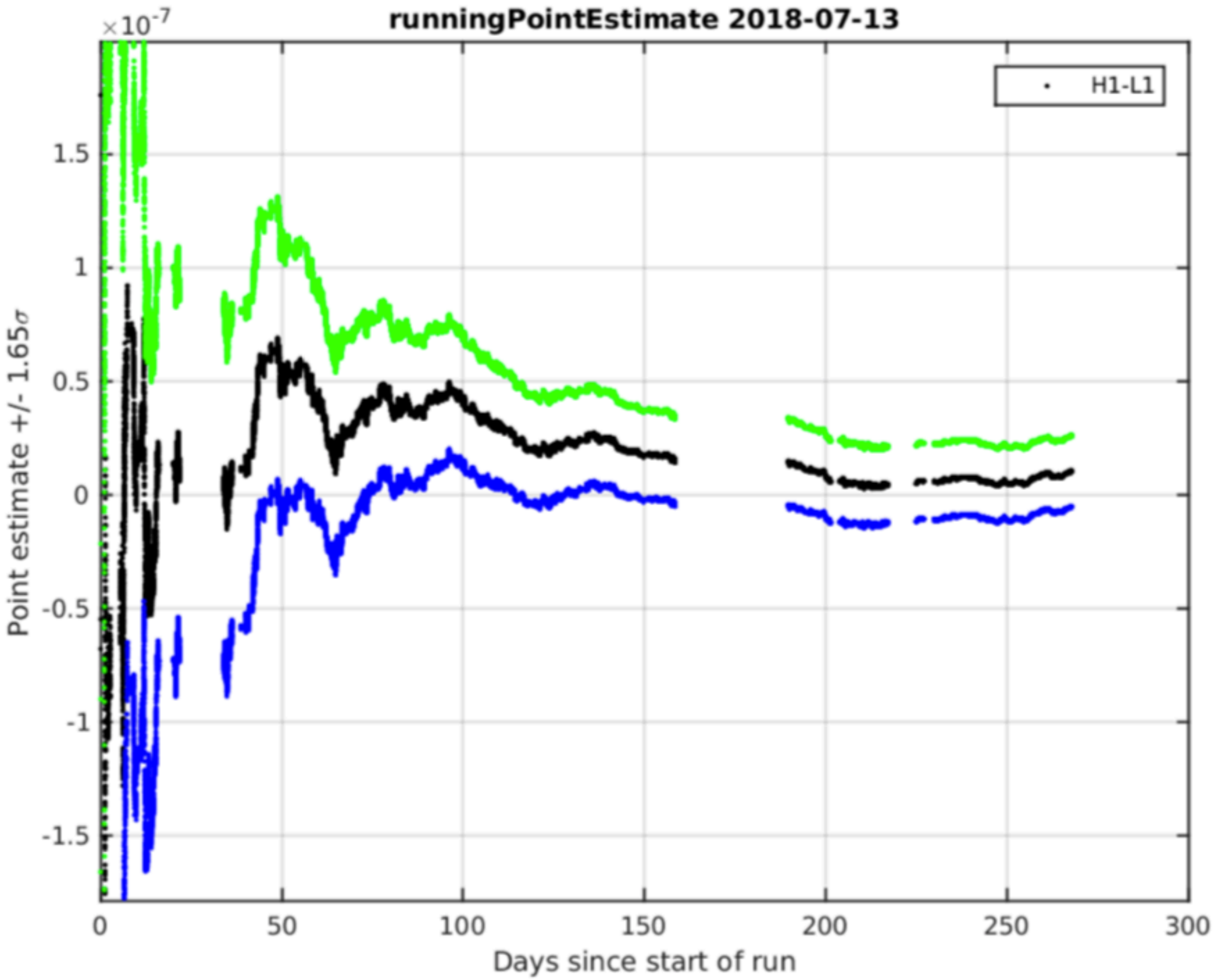
GW170817: Implications for the Stochastic Gravitational-Wave Background from Compact Binary Coalescences

The LIGO Scientific Collaboration and The Virgo Collaboration
(Dated: October 17, 2017)

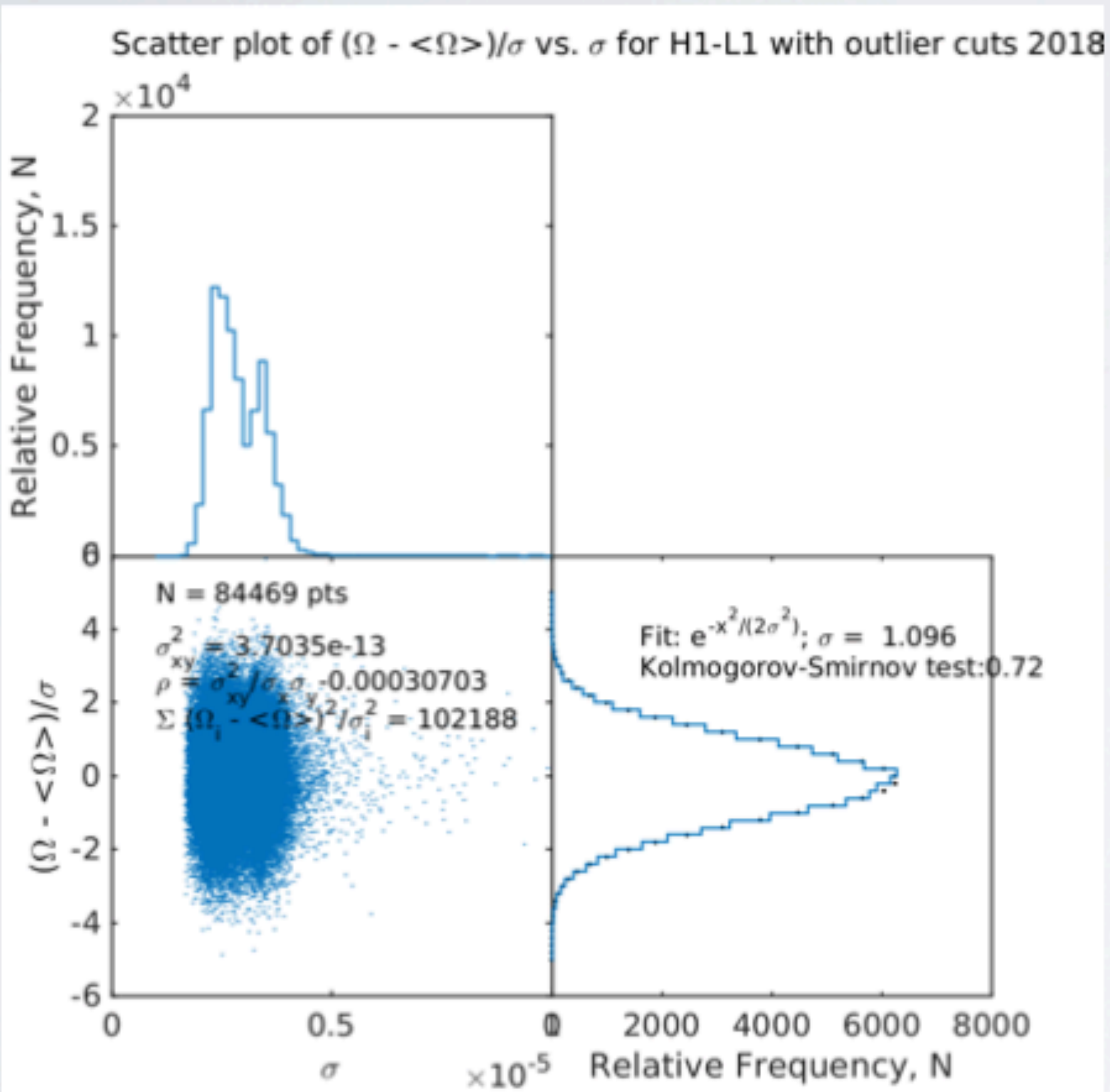
TABLE I. Estimates of the background energy density $\Omega_{\rm GW}(f)$ at 25 Hz for each of the BNS, BBH and total background contributions, along with the 90% Poisson error bounds. We also show the average time τ between events as seen by a detector in the frequency band above 10 Hz, and the number of overlapping sources at a given time λ. We quote the number given the median rate and associated Poisson error bounds.

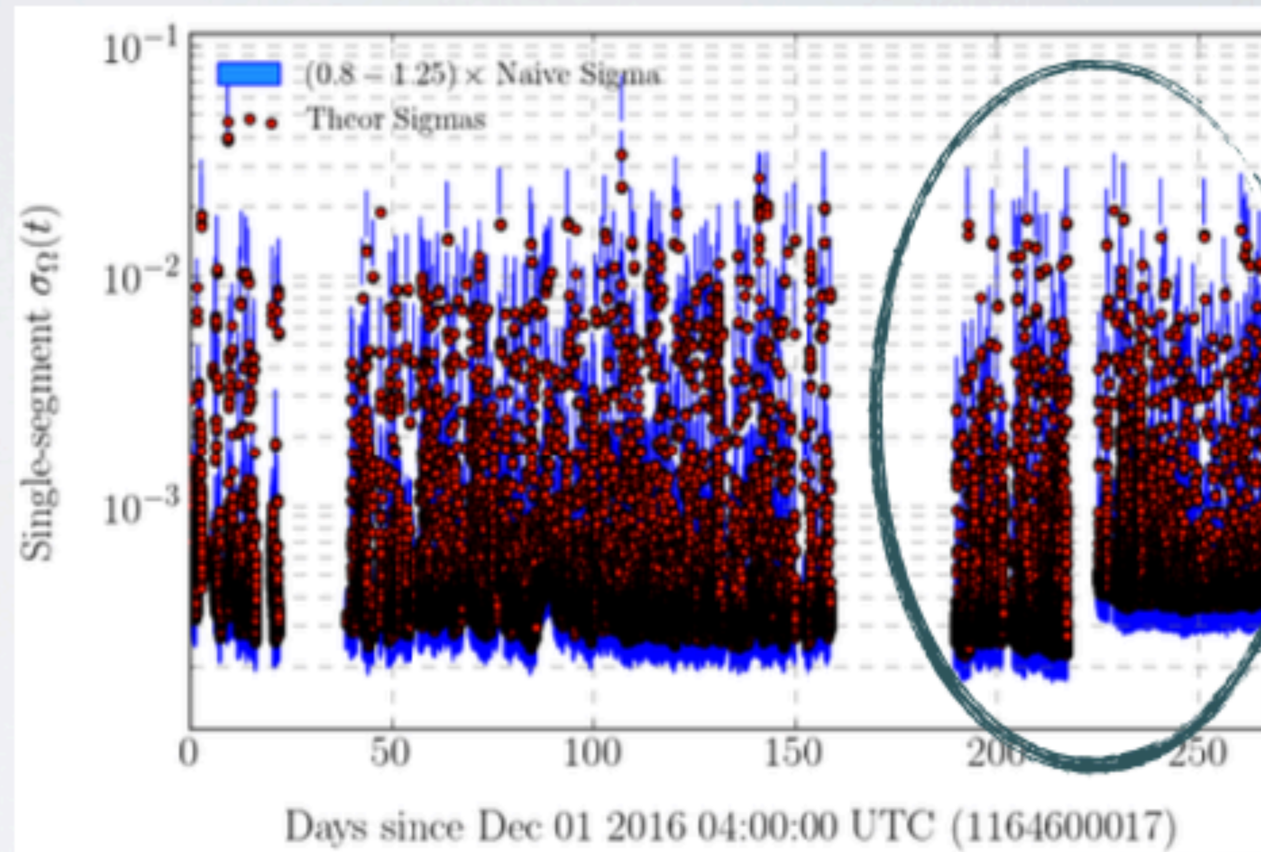
	$\Omega_{\rm GW}(25\ \text{Hz})$	τ [s]	λ
BNS	$0.7^{+1.5}_{-0.6} \times 10^{-9}$	13^{+49}_{-9}	15^{+30}_{-12}
BBH	$1.1^{+1.2}_{-0.7} \times 10^{-9}$	223^{+352}_{-115}	$0.06^{+0.06}_{-0.04}$
Total	$1.8^{+2.7}_{-1.3} \times 10^{-9}$	12^{44}_{-8}	15^{+31}_{-12}

Running Point Estimate



BIMODAL SIGMA DISTRIBUTION



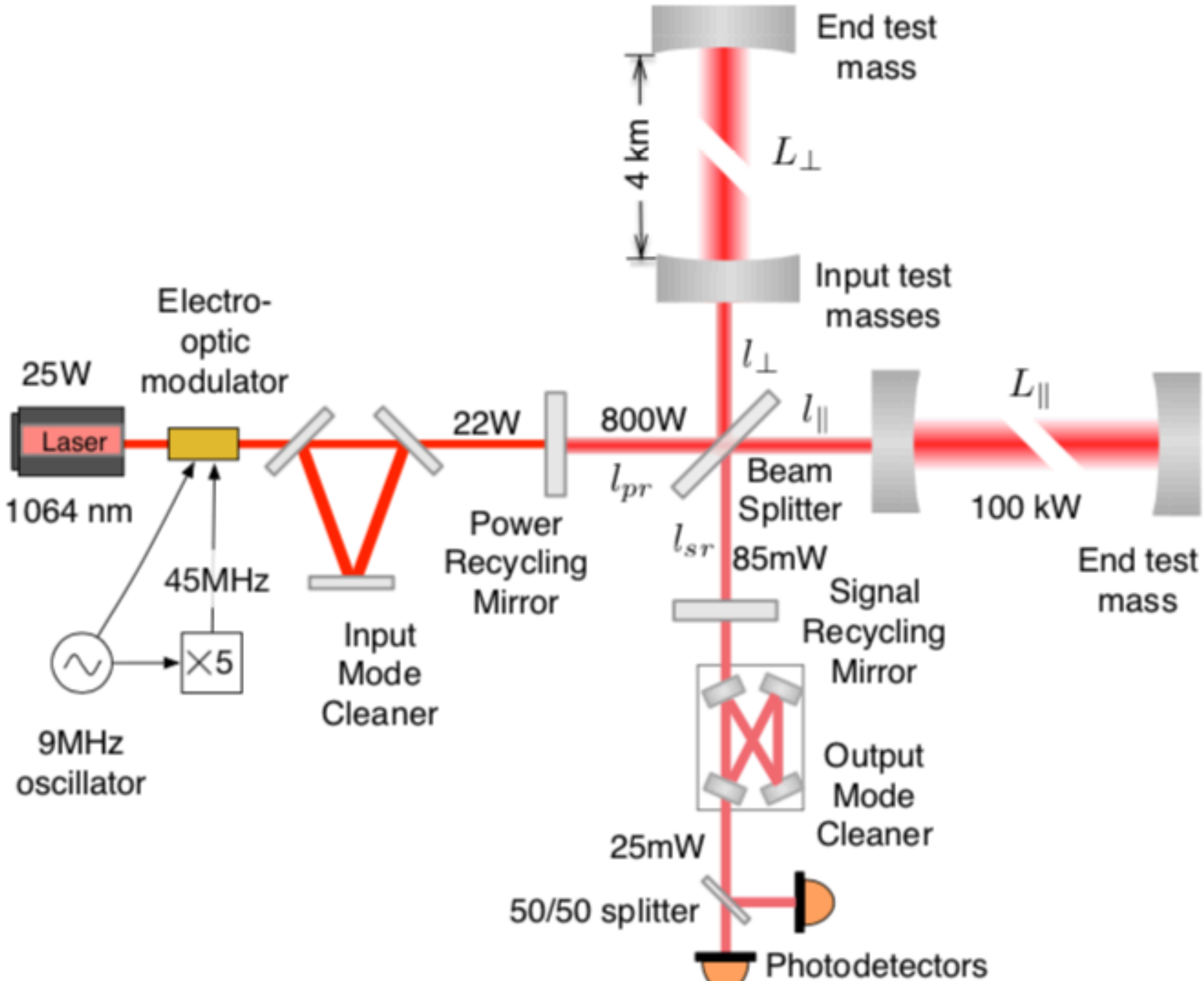


Systematic jump in sigma around July caused by earthquakes near Hanford, WA

SGWB Constraints

- COBE Multipole moments ($2 \leq \ell \leq 30$): $\Omega_{GW} h_{100}^2 < 7 \times 10^{-11} \left(\frac{H_0}{f}\right)^2$ for $H_0 < f < 30 H_0$
- Pulsar Timing Arrays: $h_c \leq 10^{-14} \longrightarrow \Omega_{GW}(f = 10^{-8}) < 10^{-8}$
- Nucleosynthesis contrains expansion rate which constrains energy density which constrains energy-density in a cosmological background of GWs: $\int \Omega_{GW}\, d(\ln f) \leq 0.07 \times (1 + z_{eq})^{-1} \approx 10^{-5}$ where $z_{eq} = 6000$

IFO Schematic



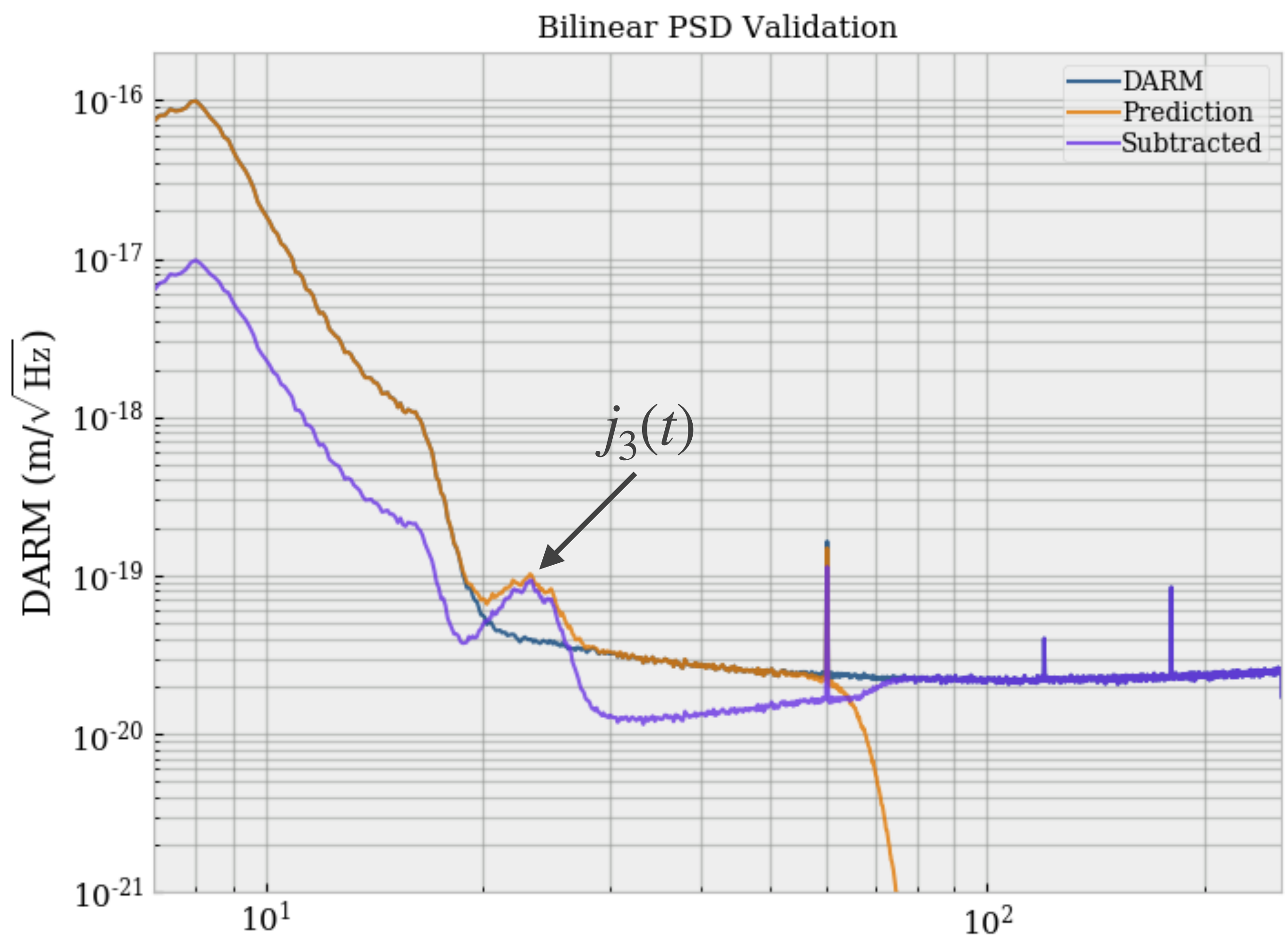
Mock Data Test

Bilinear noise added to mock DARM spectrum

Jitter Channels = $[j_1(t), j_2(t)]$

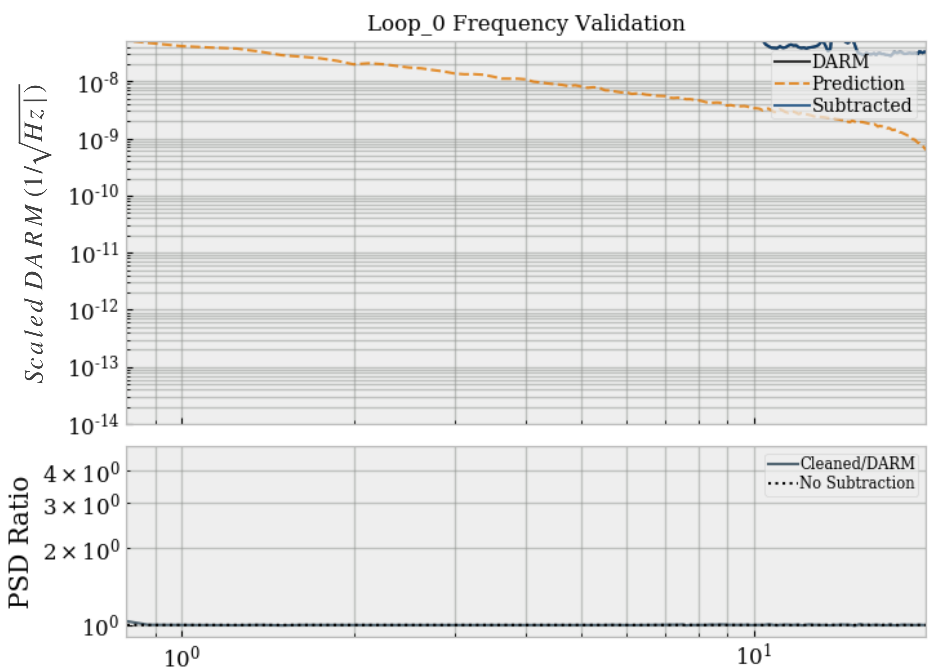
Witness Channels = $[j_1(t), j_2(t)]$

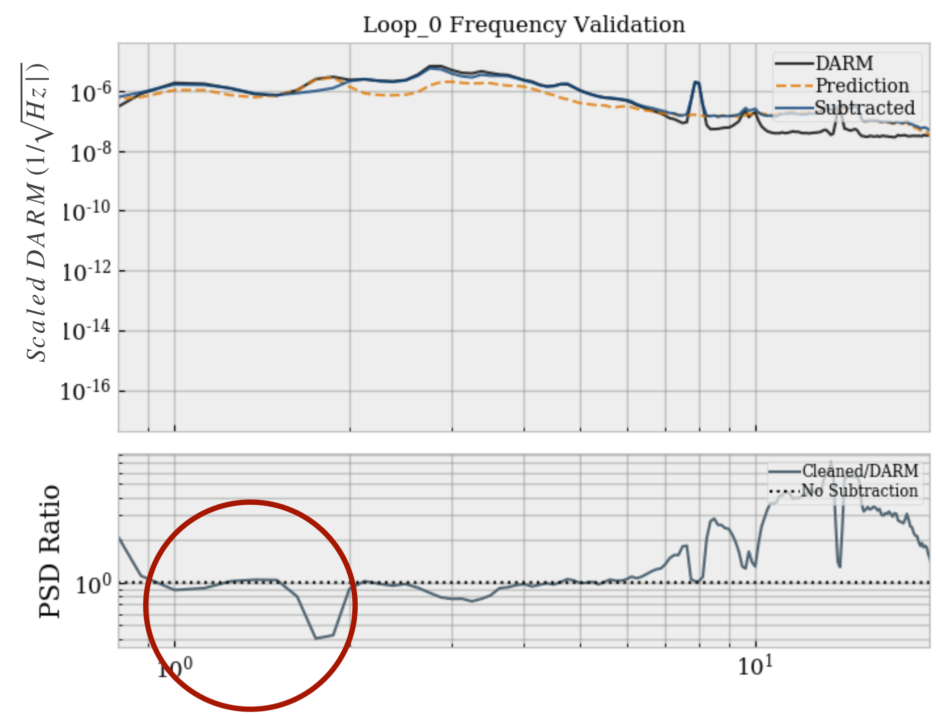
$$\text{DARM} = h(t) + \alpha * j_1(t) * j_2(t) + \beta * j_3(t)$$



Removing low frequency bilinear jitter added to mock data (generated using the MockData repo on GitLab)

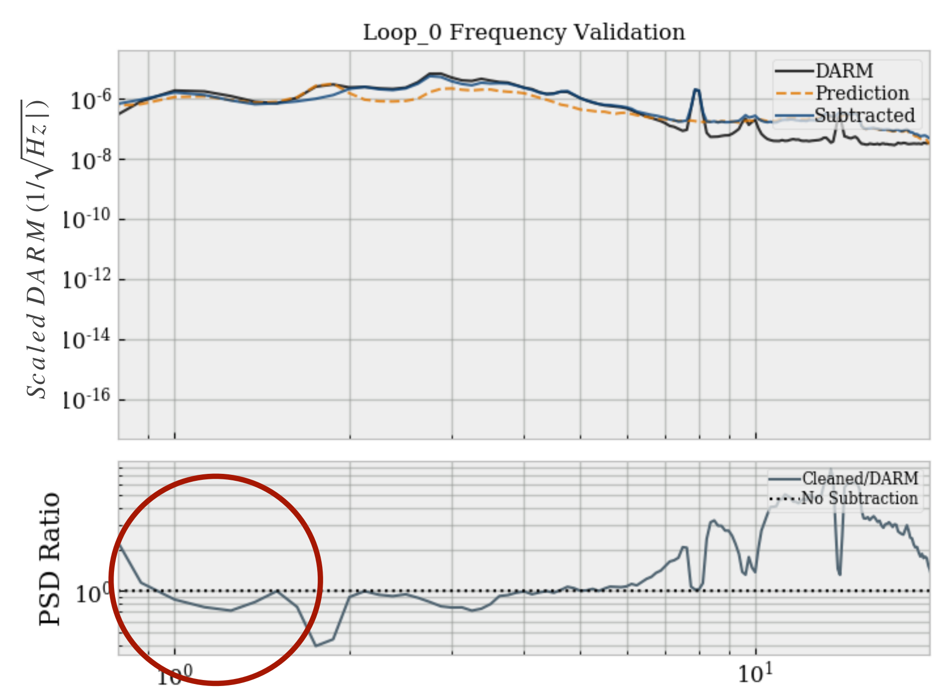
LHO O1 GW150914





Top Left: Length Sensing and Control (LSC)channels only

Top Right: Alignment Sensing and Control (ASC) channels only



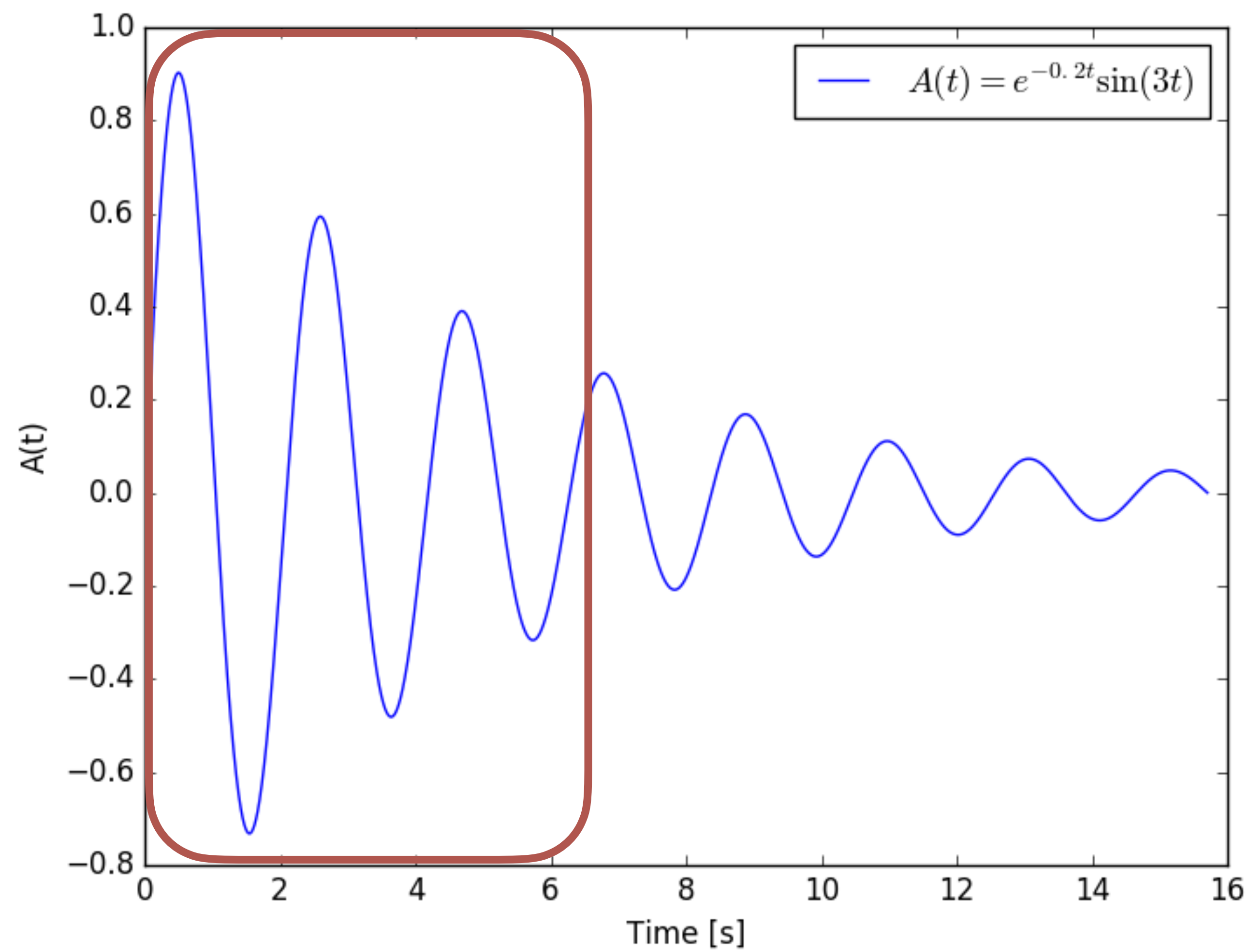
Bottom Right: LSC + ASC channels

Picking up low-frequency mode?

Recurrent Neural Networks - 1

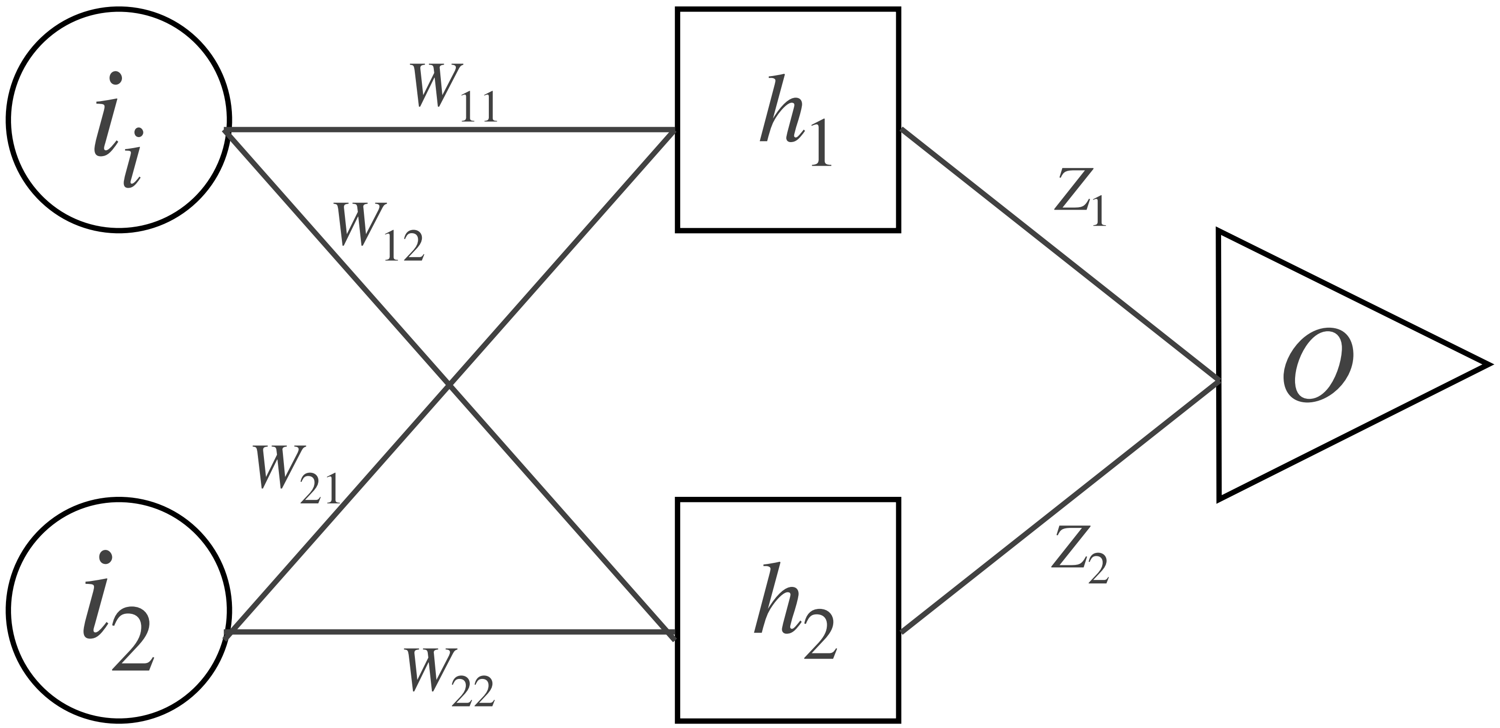
How can we capture long-term dependencies?

For the LIGO detectors, it can take seismic noise several seconds to reach the corner station and couple into the IFO, but our data is synced with GPS times. So, constantly shift data streams or try and *remember* them



Given the data in the red box, can the data that follows be predicted?

Linear Example



$$h_1 = i_1 W_{11} + i_2 W_{21}$$

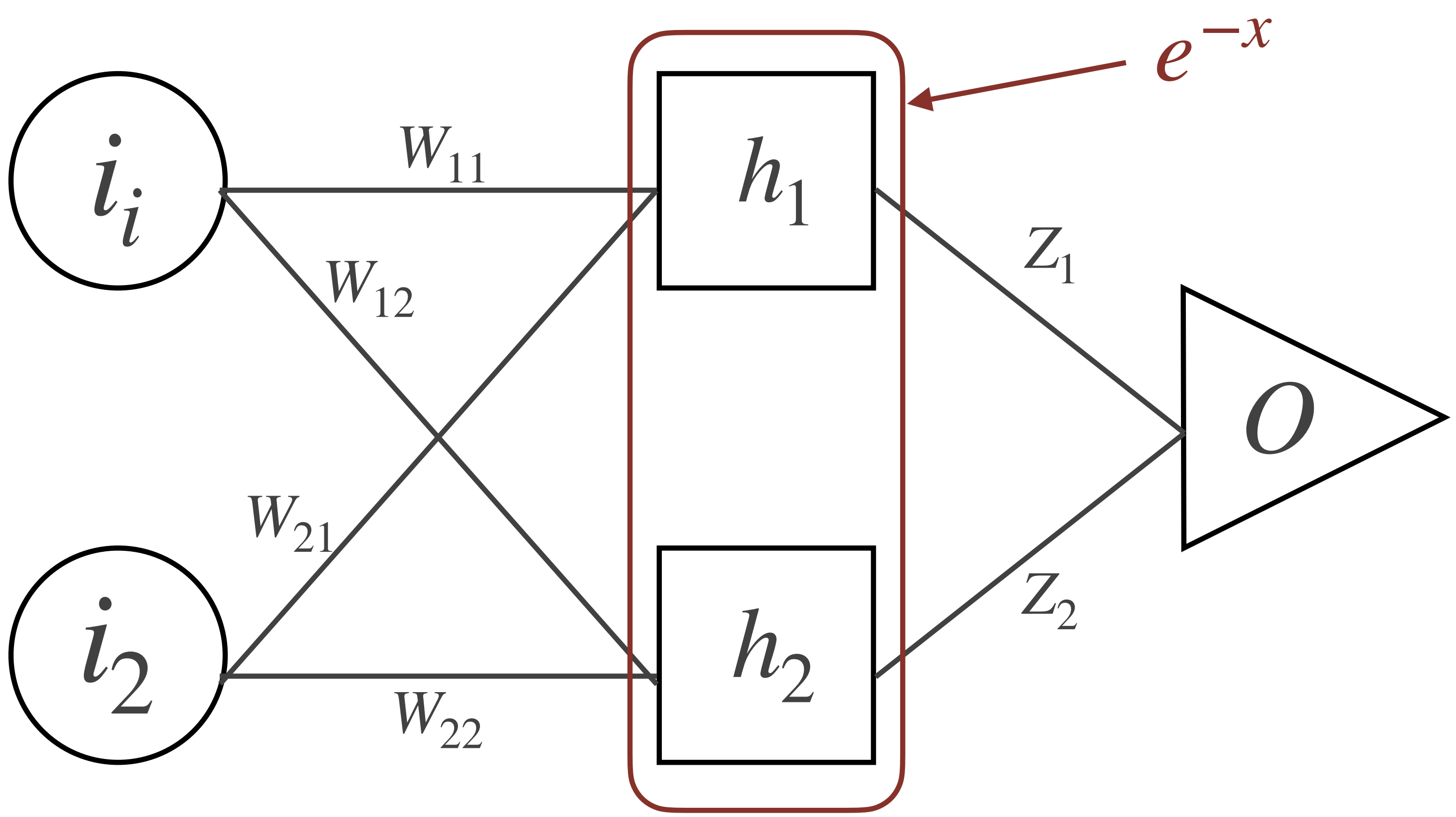
$$h_2 = i_1 W_{12} + i_2 W_{22}$$

$$O = h_1 Z_1 + h_2 Z_2$$

$$\rightarrow O = i_1(W_{11}Z_1 + W_{12}Z_2) + i_2(W_{21}Z_1 + W_{22}Z_2)$$

Output is a *linear* function of the input

Nonlinear Example



$$h_1 = e^{-(i_1 W_{11} + i_2 W_{21})}$$

$$h_2 = e^{-(i_1 W_{12} + i_2 W_{22})}$$

$$\rightarrow O = Z_1 e^{-(i_1 W_{11} + i_2 W_{21})} + Z_2 e^{-(i_1 W_{12} + i_2 W_{2s})}$$

Output is a *nonlinear* function of the input

Layers != Complexity

Layer 1:

$$y = Wx + b$$

Layer 2:

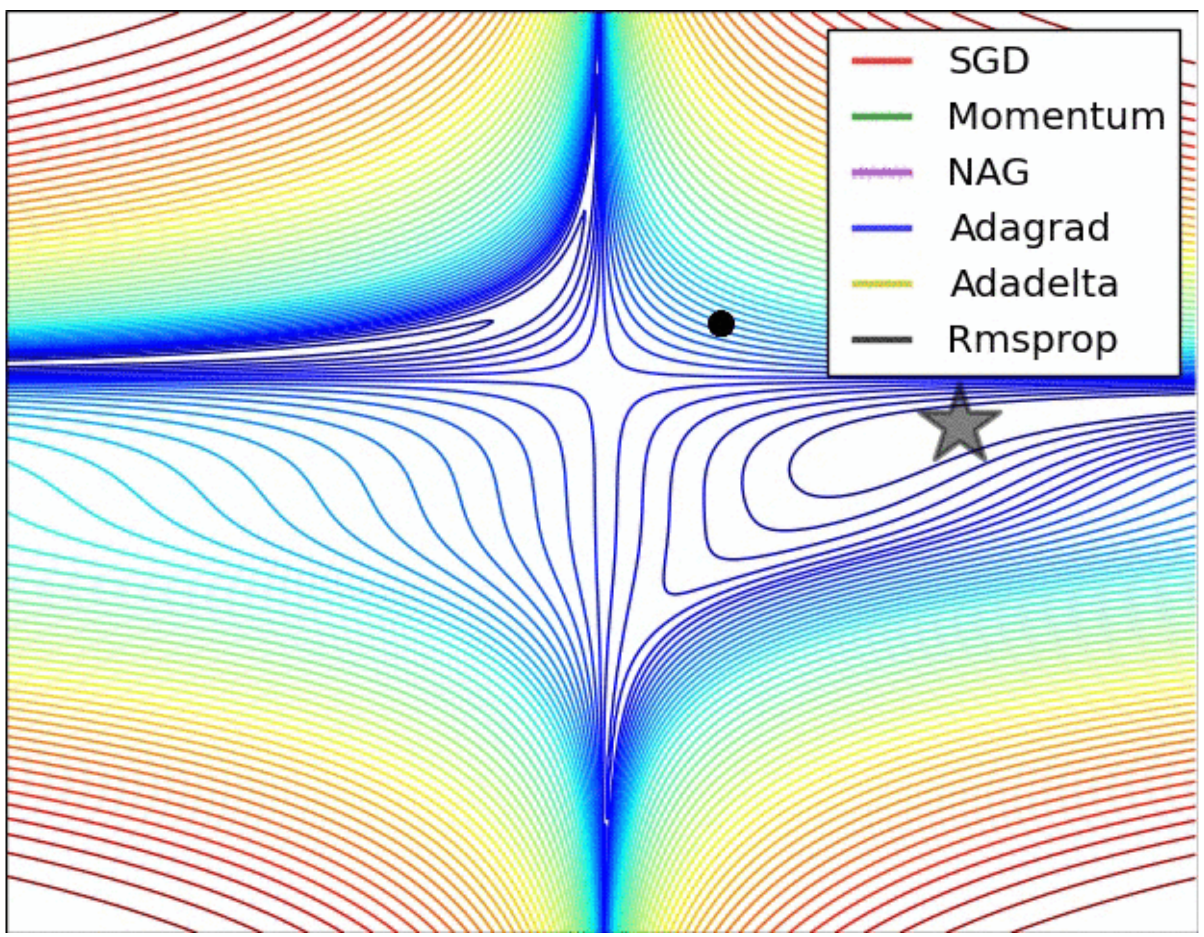
$$z = Vy + c$$

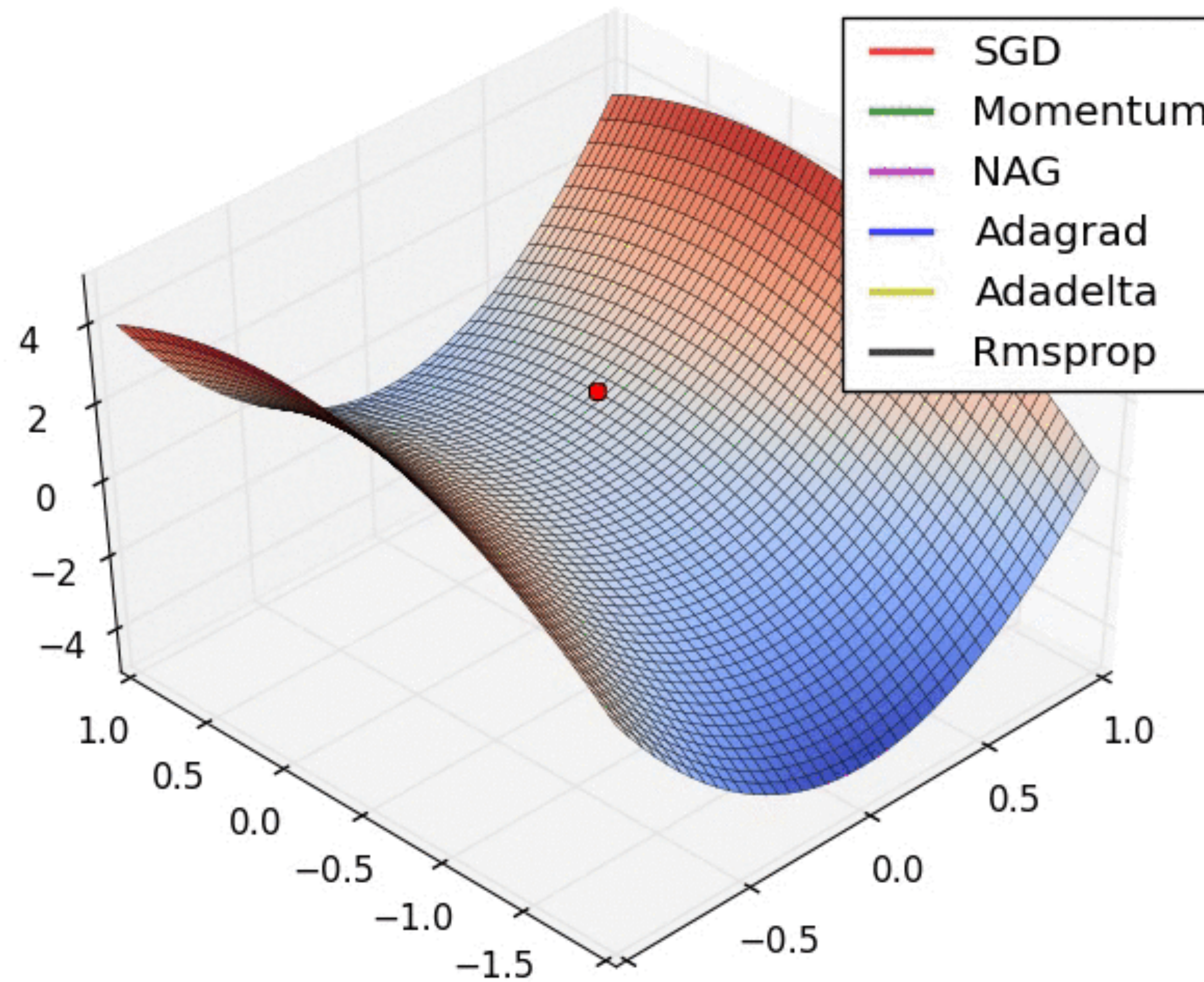
Result?

$$z = V(Wx + b) + c = W'x + (Vb + c) = W'x + b'$$

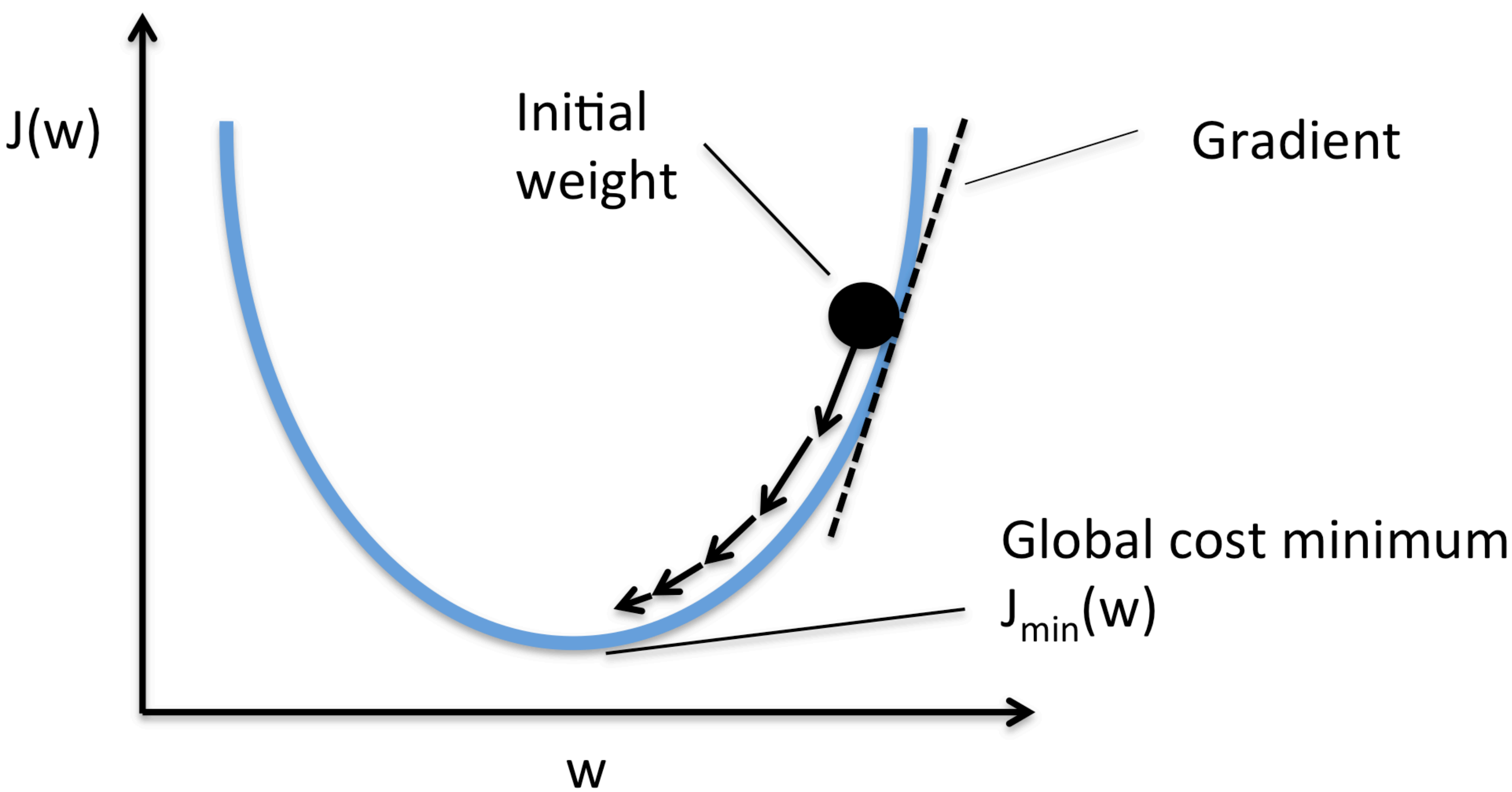
Increased training time without gaining functionality

Optimizers


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop
4
2
0
−2
−4
1.0
0.5
0.0
−0.5
−1.0
−1.5
1.0
0.5
0.0
−0.5

Gradient Descent & Cost



$$\mathbf{w} \rightarrow \mathbf{w} - \eta \vec{\nabla} \sum_i J(w_i)$$

The learning rate, eta, sets the step size as represented by the arrows in the plot. Often, the learning rate is a decaying function of the epoch to improve convergence.

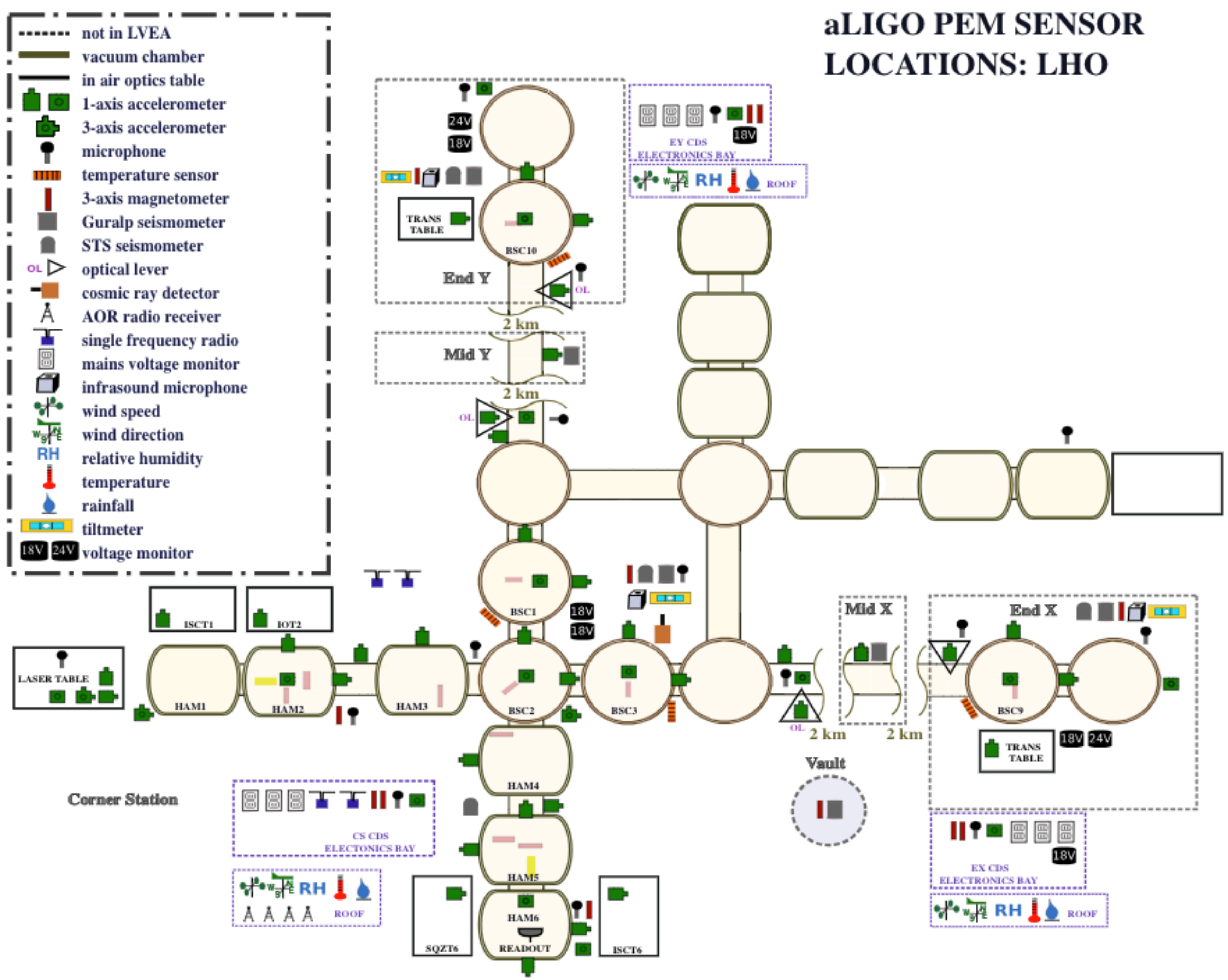
Easy to converge with simple parameter spaces. Extremely difficult and computationally costly with high dimensionality spaces.

Hanford PEM Locations

Approach - Put a large variety of physical environmental monitors everywhere

1055 channels at LHO
1000 channels at LLO

~500,000 2-channel combos
$\sim 2_{E}8$ 3-channel combos



Ref: Maggie Tse, Vincent Roma, Terra Hardwick

DeepClean Training