Application of deep-learning for gravitational wave data analysis

Amit Reza
Nikhef

Nikhef Topical Lectures on Machine Learning
June 2022

Flow of the presentation:

- **Basic of Gravitational wave astronomy**
  - What is Gravitational waves?
  - Discovery of Gravitational waves
  - Sources of Gravitational waves
- **Basic of Gravitational wave data analysis**
  - Matched Filtering scheme
  - Signal Consistency test
  - Parameter estimation
- **Application of machine learning for Gravitational wave data**
  - Binary classification problem
  - Multi-label classification problem
  - Regression problem

What are Gravitational Waves?? How can we observe them??

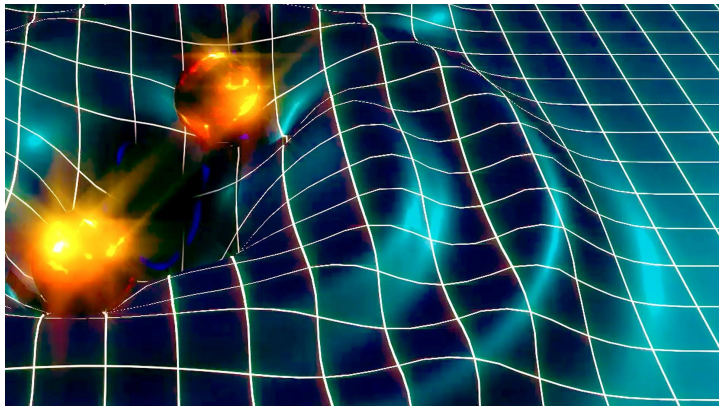
Curvature of space

Distribution of mass/energy

$$G_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta}$$

Some constants

GW treated as a phenomenon resulting from the curvature of spacetime.



- **Direct evidence for a phenomenon that was first predicted 100 years ago by Einstein**

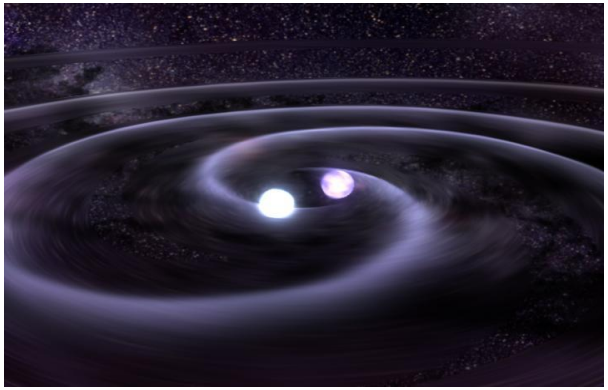
Gravitational waves are ripples in Space-Time fabric

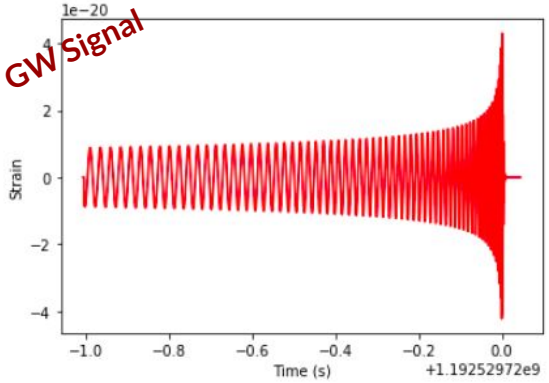
They are 'waves' of distorted space that radiate from the massive accelerating objects (such as neutron stars or black holes orbiting each other)

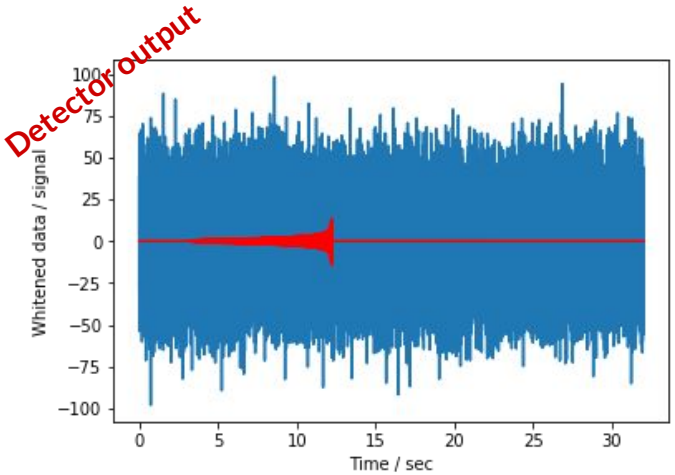
Observation and Detection : Very faint signals coming from sources billions of light years away; can only be detected through very sensitive Interferometers(LIGO, Virgo, Indigo, etc)

Gravitational wave searches from compact binary coalescence

How to extract a faint signal buried in a noise?







$$\vec{d} = \vec{s} + \vec{n}$$

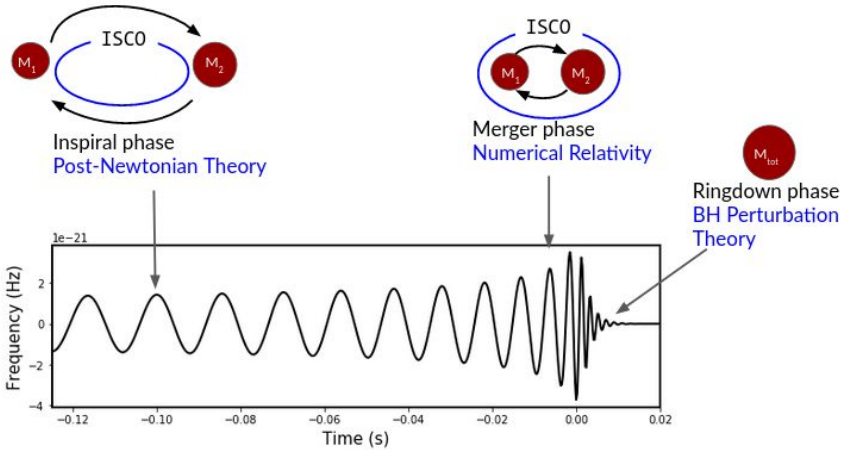
Modeled searches

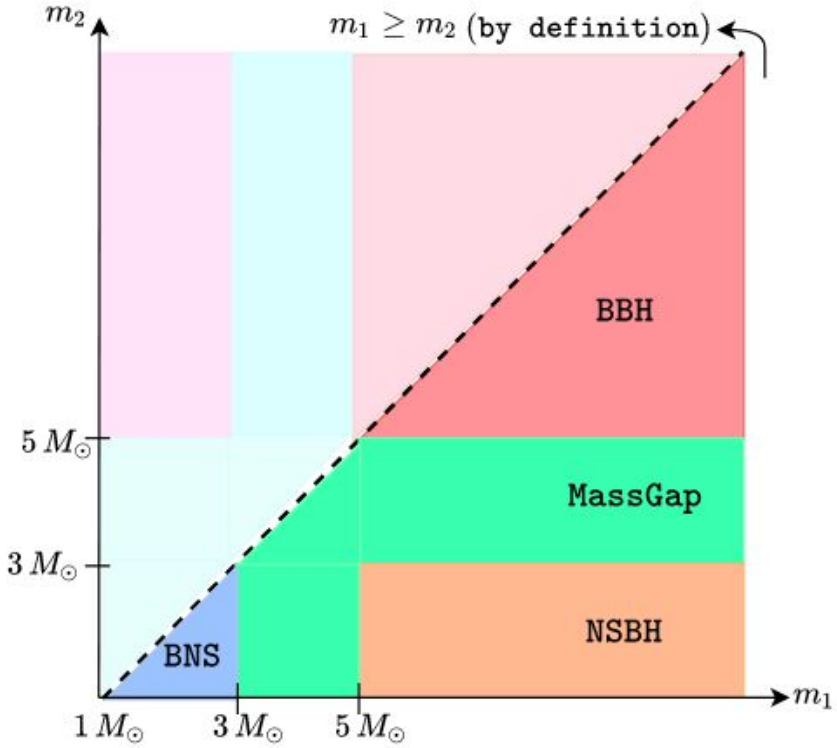
- Search for known signal shape in detector noise.
- Signal parameters are unknown.

Nearly real time analysis of the data is required.

Sources of GW

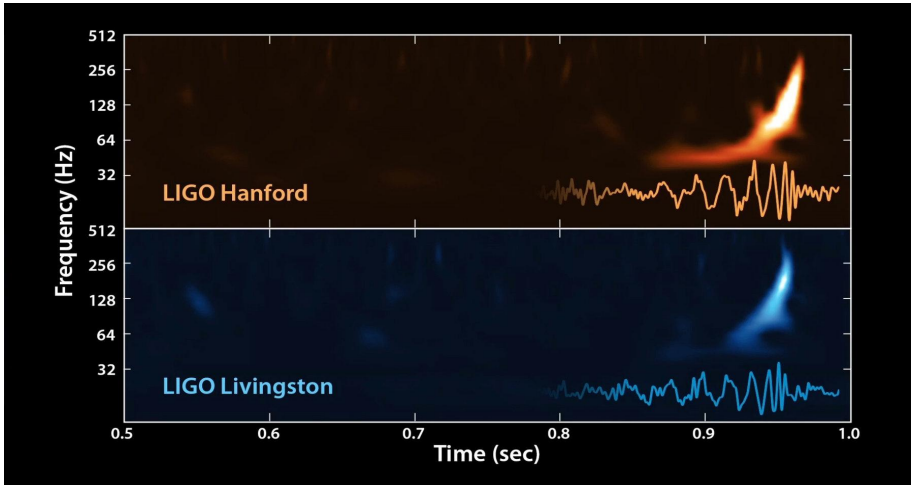
- Compact Binary Coalescence
  - BBH
  - BNS
  - NSBH
- Burst
  - Core-collapse Supernova
- Continuous waves
  - Rotating NSs
- Stochastic Background
  - Primordial Background
  - Superposition of unresolved GW sources





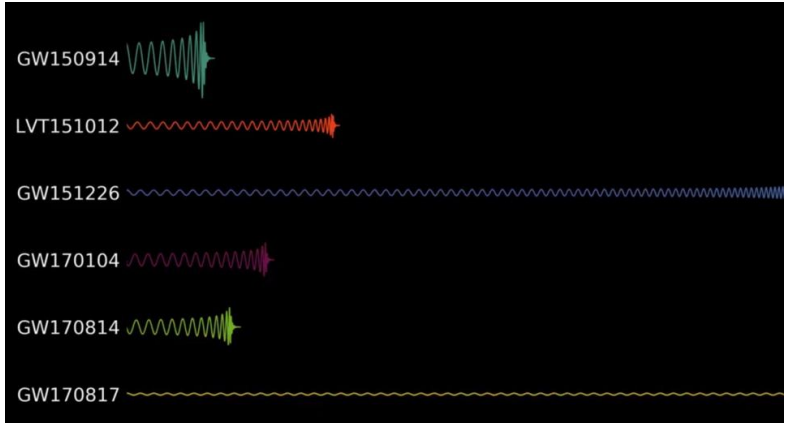
LIGO Detector & First Observation!

Hanford Observatory



First Detection (14th Sep, 2015)

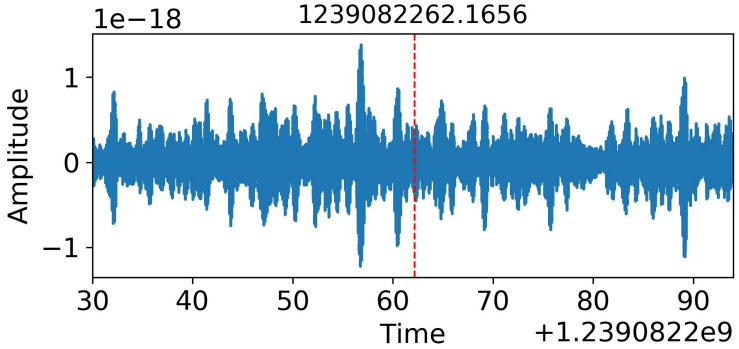
Other detections



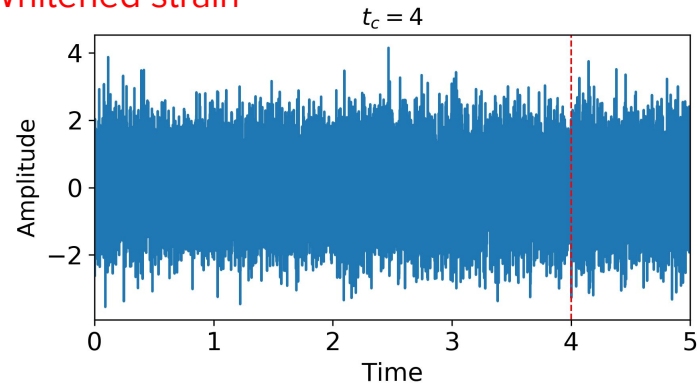
- Hanford (USA)
- Livingston (USA)
- Virgo (Italy)
- KAGRA (Japan)

Time-frequency map of strain data

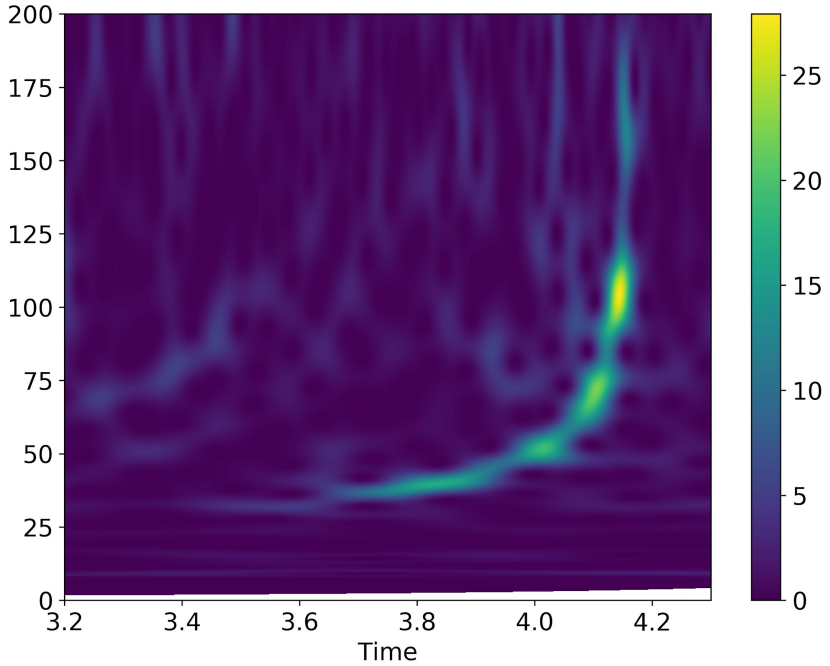
L1 strain surrounding GW190412

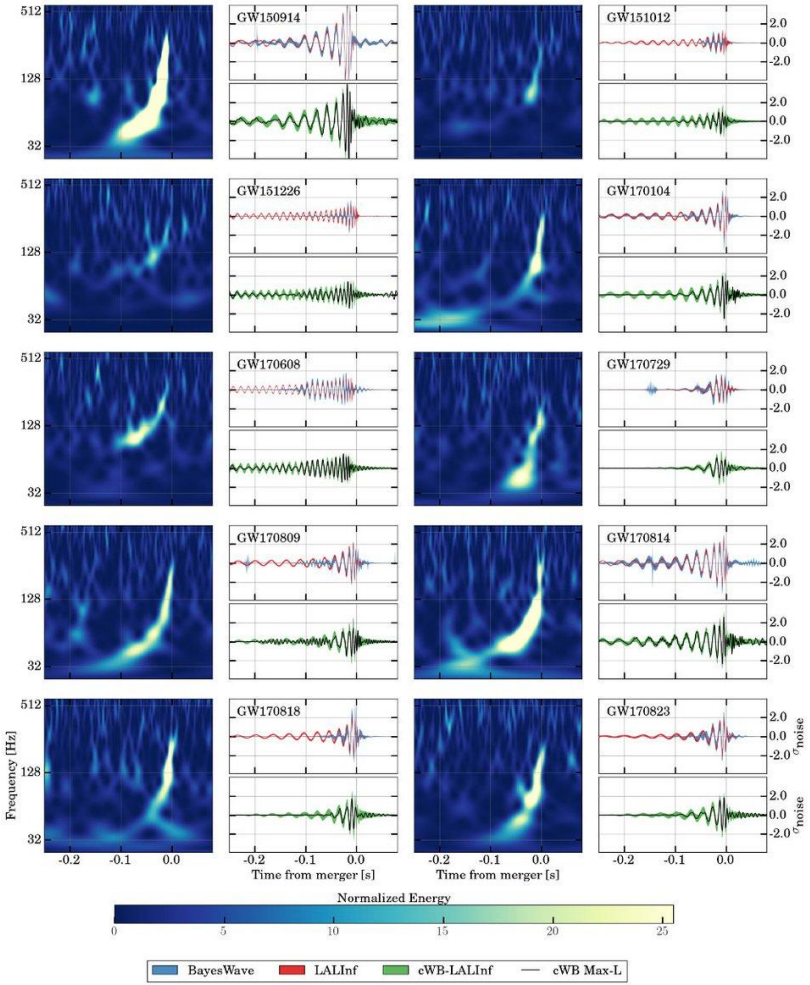


Whitened strain



Time-frequency map of whitened strain data





GW150914: First Detection (14th Sep, 2015)

GW150914, solved the following unsolved puzzles:

- ❏ It provided direct evidence of the existence of black holes.
- ❏ It validated the existence of a binary black hole system.
- ❏ It also validated the presence of massive black holes heavier than 25 solar masses.
- ❏ It validated the Einstein's theory of General Relativity.

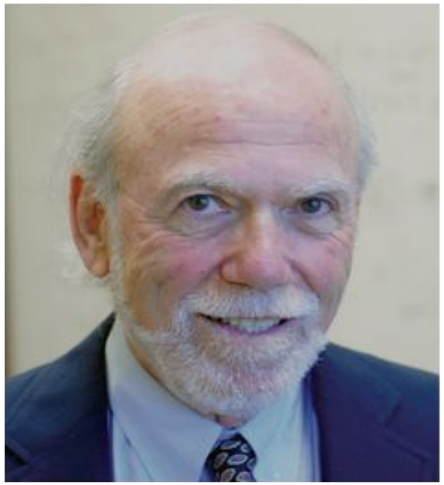
Who Are they?

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2017

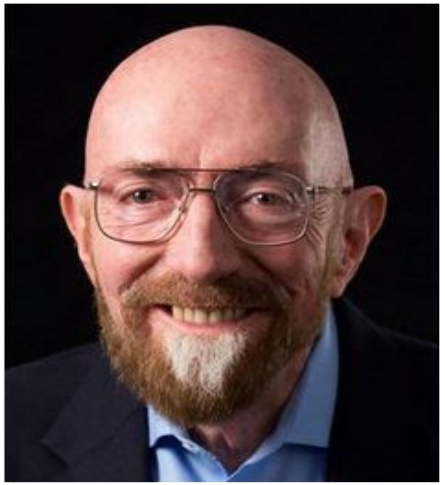
With one half to

Rainer Weiss
MIT

and the other half jointly to



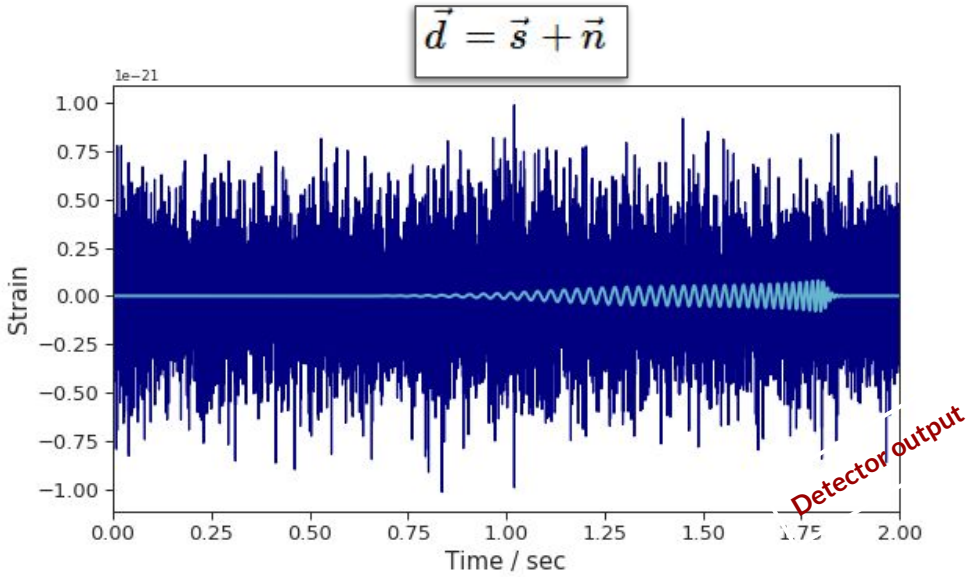
Barry C. Barish
CalTech



Kip S. Thorne
CalTech

"for decisive contributions to the LIGO detector and the observation of gravitational waves."

Gravitational wave searches from compact binary coalescence (CBC)



Real time analysis of the data is required.

1. Enabling faster and more efficient follow up of EM counterpart.
2. Potential Probe for Multi messenger astronomy.

Modeled searches

- Search for known signal shape in detector noise.
- Signal parameters are unknown.

Existing CBC search pipelines

- PyCBC
- GstLAL
- MBTA
- SPIIR

Low-latency search pipelines for CBC sources

1. PyCBC-Live [Frequency domain based matched filtering]
2. GstLAL [Basis based time-domain matched filtering, basis computed using SVD]
3. SPIIR [Series of infinite impulse response filters]
4. MBTA online [Multi-rate filtering]

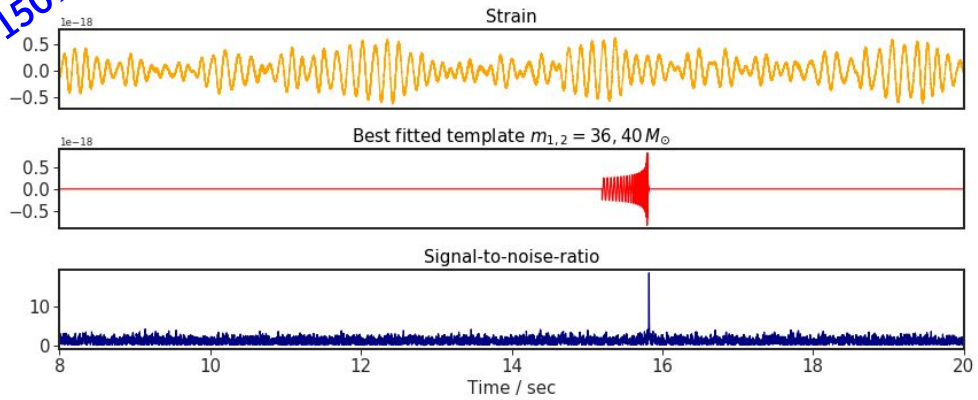
$$\rho_\alpha(\Delta t) = \vec{h}^{\alpha}(\Delta t) \cdot \vec{d}$$

⇩ SNR

Analytical waveform

Detector-output

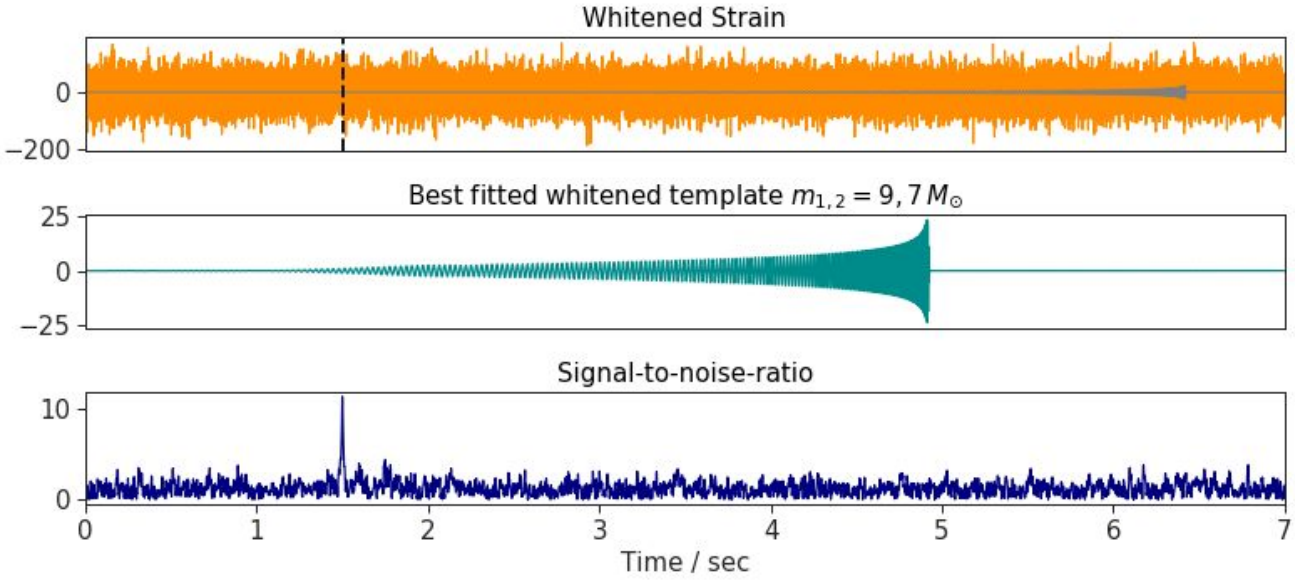
GW150914



First Step [Common to all pipelines]: Matched filtering between detector output (data) and a set of analytical waveforms.

Matched Filtering scheme

$$\left\langle d|h^{\alpha}\right\rangle\Big|_{\Delta t} = 4\,\mathbf{Re}\left[\int_{\mathtt{flow}}^{\mathtt{fhigh}} \underbrace{\frac{\tilde{d}(f)}{\sqrt{S_h(f)}}}_{\substack{\mathtt{Whitened}\\ \mathtt{data}}} \underbrace{\frac{\tilde{h}^{\alpha}(f,\vec{\lambda})}{\sqrt{S_h(f)}}}_{\substack{\mathtt{Whitened}\\ \mathtt{template}}} e^{2i\pi f\Delta t}\, df\right]$$



Matched Filtering Scheme

a. Generate a set of analytical waveforms.

b. Calculate the correlation of the analytical waveforms with the data.

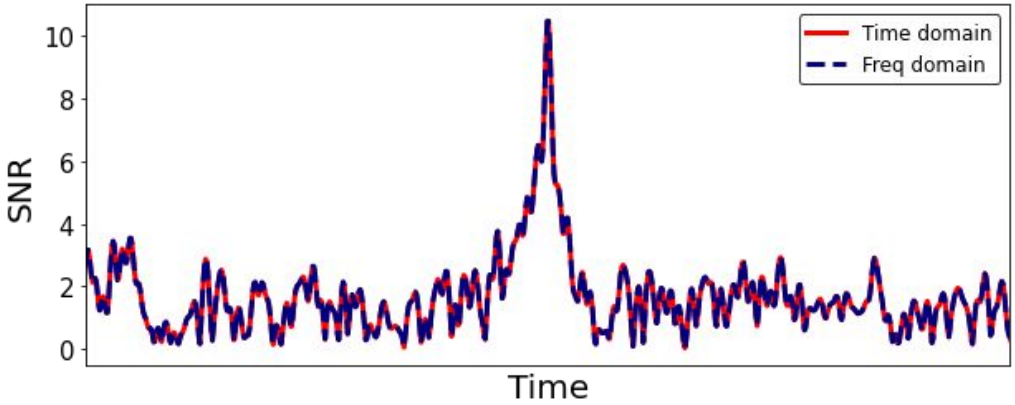
The collection of analytical waveforms known as template bank.

$$\mathtt{Template\ Bank} = \{h^{\alpha}(\vec{\lambda}) : \alpha = 1, 2, \cdots, N_T\}$$

Matched Filtering in Time-domain

Convolution

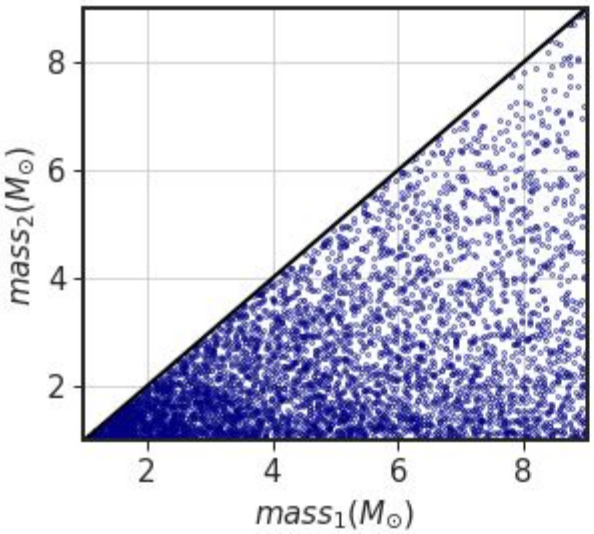
$$\rho(\Delta t) = \langle s, h \rangle|_{\Delta t}$$
$$= C(\hat{h}) \cdot \hat{s}$$

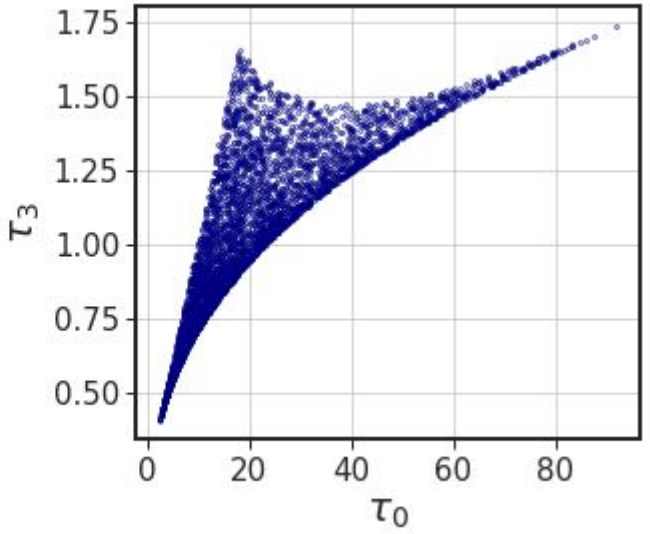


$$C(\hat{h}) = \begin{pmatrix} \hat{h}_0 & \hat{h}_1 & \hat{h}_2 & \cdots & \hat{h}_{N_s-1} \\ \hat{h}_{N_s-1} & \hat{h}_0 & \hat{h}_1 & \hat{h}_2 & \cdots \\ \hat{h}_{N_s-2} & \hat{h}_{N_s-1} & \hat{h}_1 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ \hat{h}_1 & \hat{h}_2 & \cdots & \hat{h}_{N_s-1} & \hat{h}_0 \end{pmatrix}$$

$$\theta = \begin{bmatrix} \alpha & \text{right ascension} \\ \beta & \text{declination} \\ r & \text{distance} \\ t_a & \text{arrival time at geocenter} \\ \iota & \text{inclination angle} \\ \psi & \text{polarization angle} \\ \phi_c & \text{coalescence phase} \\ D & \text{luminosity distance} \\ \hline m_1 & \text{first mass component} \\ m_2 & \text{second mass component} \\ \vec{\mathbf{S}}_1 & \text{first spin component} \\ \vec{\mathbf{S}}_2 & \text{second spin component} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \text{extrinsic parameters} \\ \\ \\ \\ \text{intrinsic parameters} \end{matrix}$$

Representation of Template Bank





Template placement algorithms:

- Stochastic
- Geometric
- Hybrid

$$\tau_0 = \frac{5}{26}\frac{1}{\pi f_0 \eta}(\pi M f_0)^{-5/3},$$

$$\tau_3 = \frac{1}{8}\frac{1}{f_0 \eta}(\pi M f_0)^{-2/3},$$

$$\eta = \frac{m_1 m_2}{M^2}$$

Computational Cost of Matched Filtering

aLIGO-Era

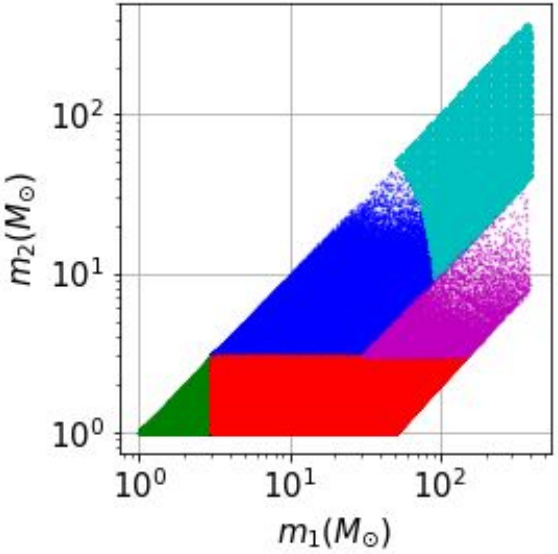
Matched Filtering Cost (per template) : $N_s \log N_s$

N_s = Time Duration × Sampling frequency

For example : $N_s = 10^6$

Matched Filter Cost per template : $\mathcal{O}(10^8)$

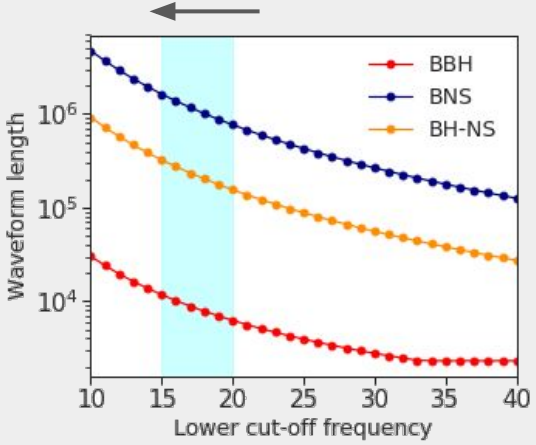
O_3 Template bank (GstLAL)



Pipeline	Observation	Search range	#Templates
GstLAL	O_1	$2M_\odot \le M \le 100M_\odot$	6.5×10^4
	O_2	$2M_\odot \le M \le 400M_\odot$	6.77×10^5
	O_3	$2M_\odot \le M \le 400M_\odot$	6.77×10^5
PyCBC	O_1	$2M_\odot \le M \le 100M_\odot$	7.5×10^4
	O_2	$2M_\odot \le M \le 500M_\odot$	4×10^5
	O_3	$2M_\odot \le M \le 500M_\odot$	4.1×10^5

If # templates : 10^5

Matched Filter Cost (for whole bank) : $\mathcal{O}(10^{13})$

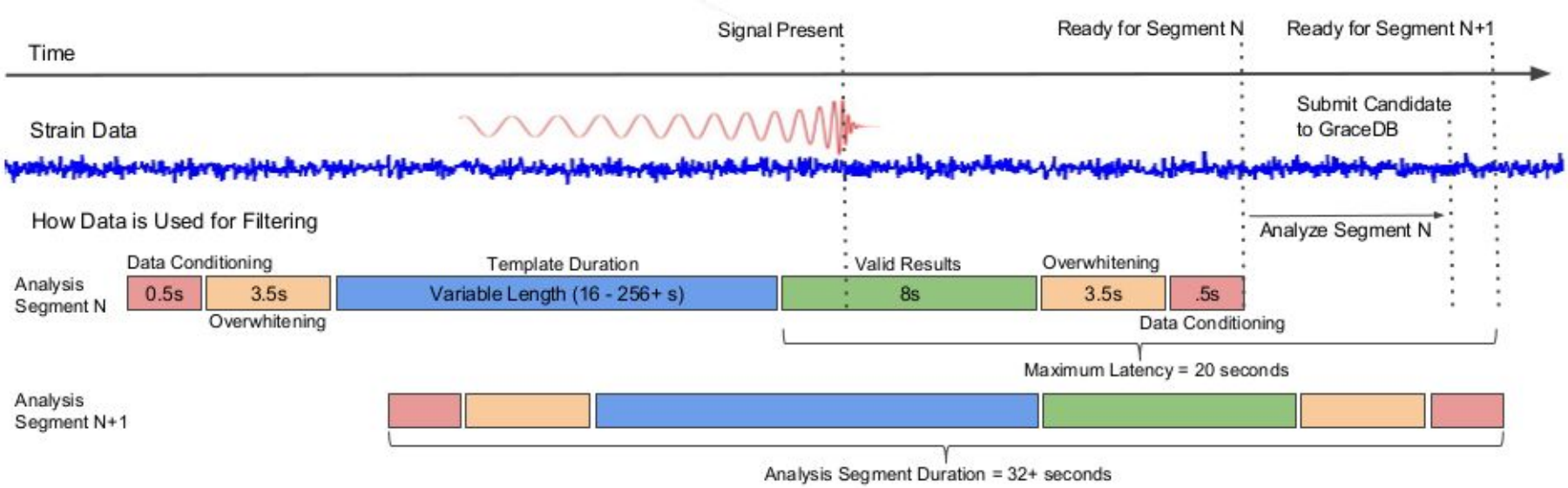


Challenges in CBC searches in advanced LIGO era

- Explosion of a large number of templates.
- Longer Templates

Scalable and efficient search methods are required.

PyCBC Live



- The average latency of identification of any GW event is 16 sec, and a maximum latency in this pipeline is 20 sec.
- The low latency occurs because several steps are involved in this analysis, starting from the data conditioning to matched filtering.
- In practice, to obtain crucial information such as sky localization; it also required at least 10 sec.

PyCBC workflow

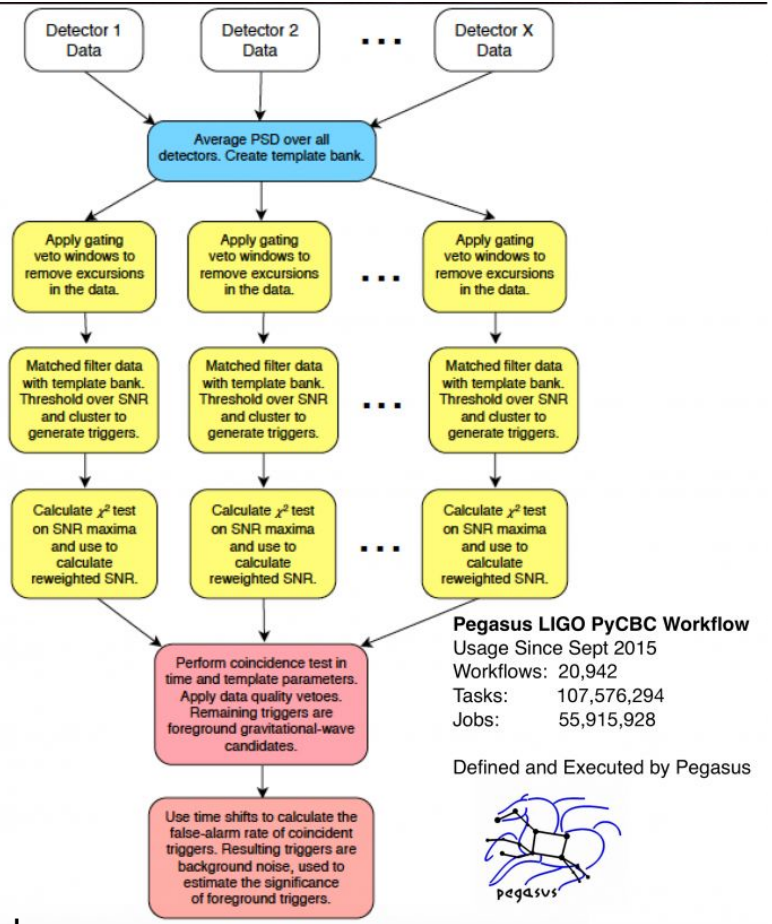


Image Credit: Samantha Usman, Duncan Brown et al

- The data collected from the GW observatories to the GW event identification in the GraceDB through the GW data analysis using PyCBC-Live low-latency pipeline.
- Several work nodes process different regions of the template bank, and the control node collects all the triggers and identifies significant candidate events.
- An alert to the astronomers from the Gamma-ray Coordinates Network (GCN).

Core design of GstLAL pipeline

Template bank $\{\vec{h}^{\alpha}\}$

Template matrix **H**

SVD

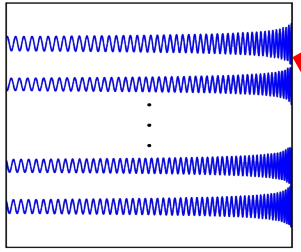
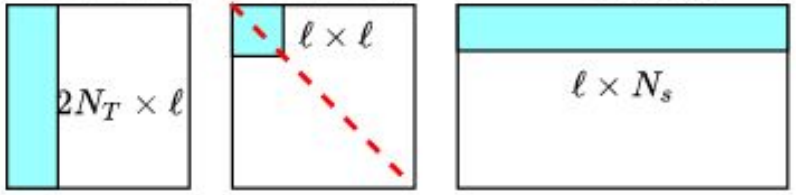
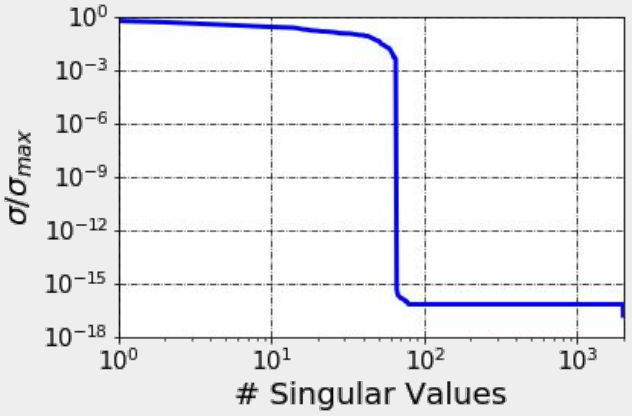


Fig Credit: Cannon et. al. (2010).

$\mathbf{U}_{2N_T \times 2N_T}$ $\mathbf{\Sigma}_{2N_T \times 2N_T}$ $\mathbf{V}^T_{2N_T \times N_s}$

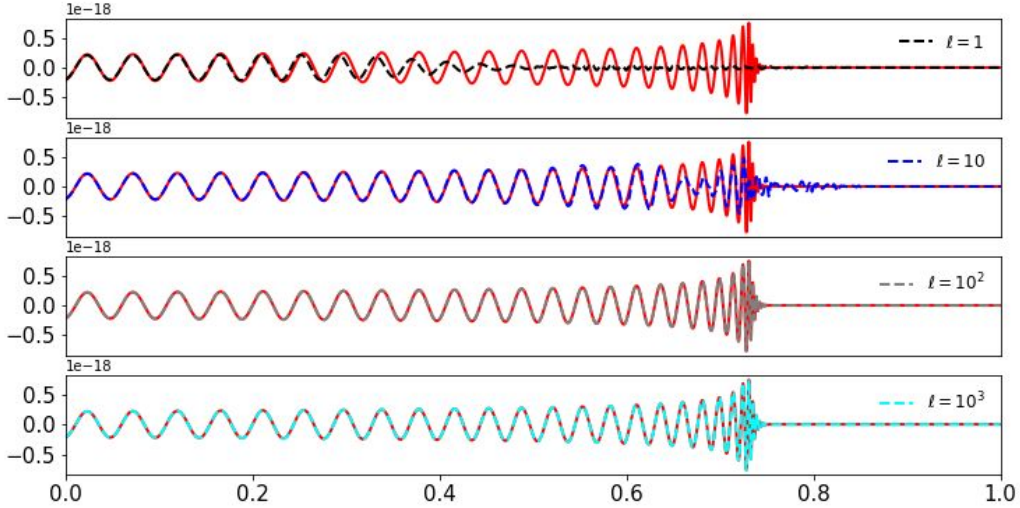


Left Basis | Singular values | Right Basis



Truncation on basis vectors can be done based on singular value spectrum.

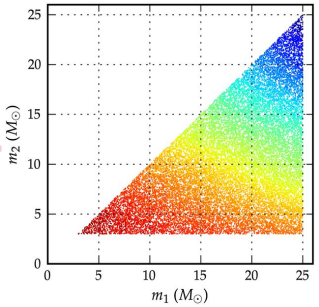
Reconstructed waveform



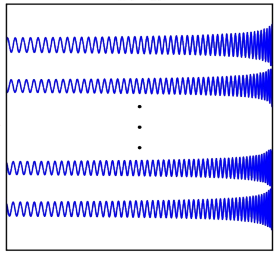
Control on SNR Reconstruction Accuracy

$$\left\langle \frac{\delta\rho}{\rho} \right\rangle = \frac{1}{4N_T} \sum_{\nu=\ell+1}^{2N_T} \sigma_\nu^2$$

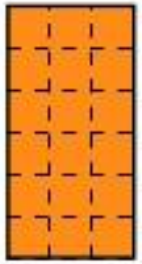
Basic steps in GstLAL-framework



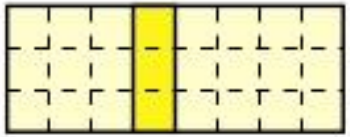
Step-1



Step-2



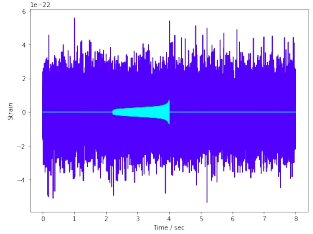
SVD decomposition

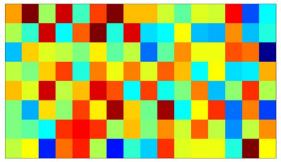


Step-3

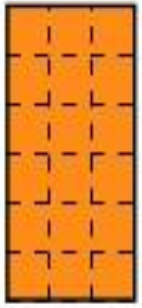
Circulant

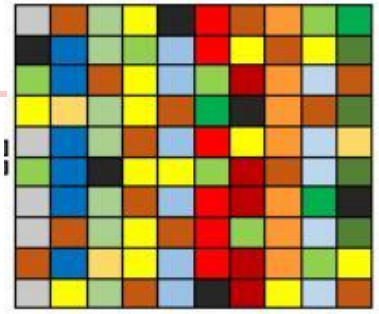
Step-4



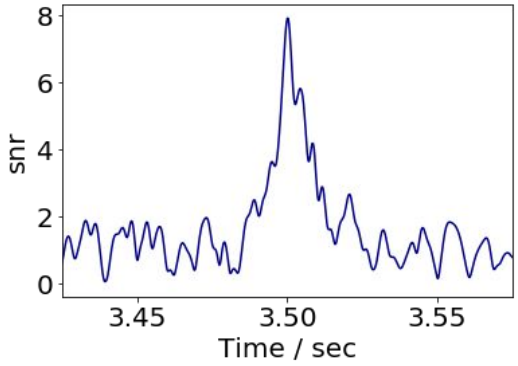


Step-5





Step-6



Collection of coincidence trigger over multiple detectors network

If triggers from the same template are within the GW propagation time between observatories, the collection of triggers is called coincident triggers.

- Given a trigger in one detector, we check for corresponding triggers in the other detector within an appropriate time window.
- It takes into account the maximum GW travel time between detectors and statistical fluctuations in the measured event time due to detector noise.
- For the two LIGO detectors, the time window is typically ±15 ms.

- The threshold in each detector for identifying a trigger is very low.
- Billions of triggers based on SNR threshold.
- Tens of millions triggers found in coincidence event.
- Required statistics for event ranking.

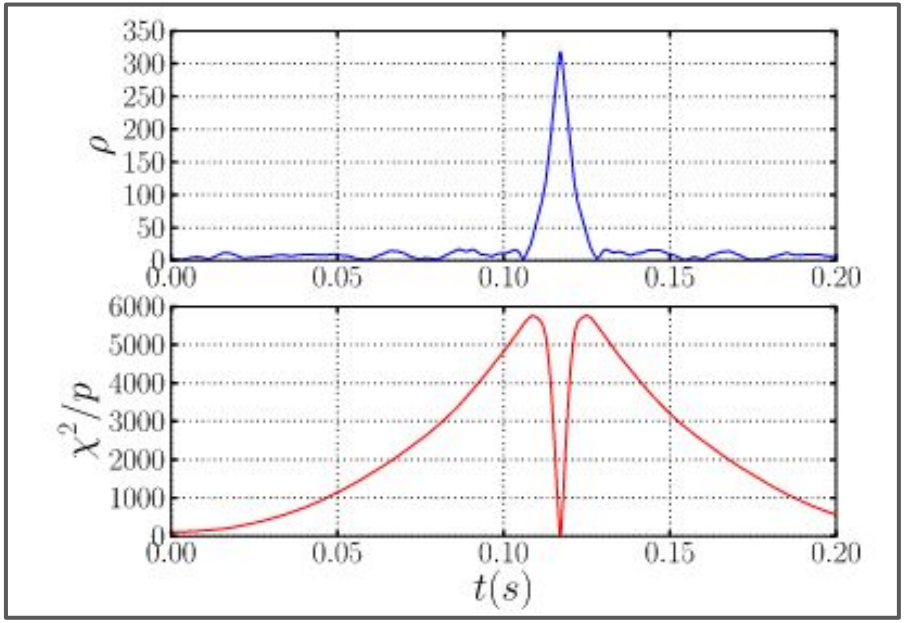
Signal consistency Test: Chi-square test [B.Allen et al. 2004, S. Babak et al. 2012]

- For a GW signal that matches the template waveform exactly, the sum of squared residuals follows the chi-square distribution with 2p−2 degrees of freedom.

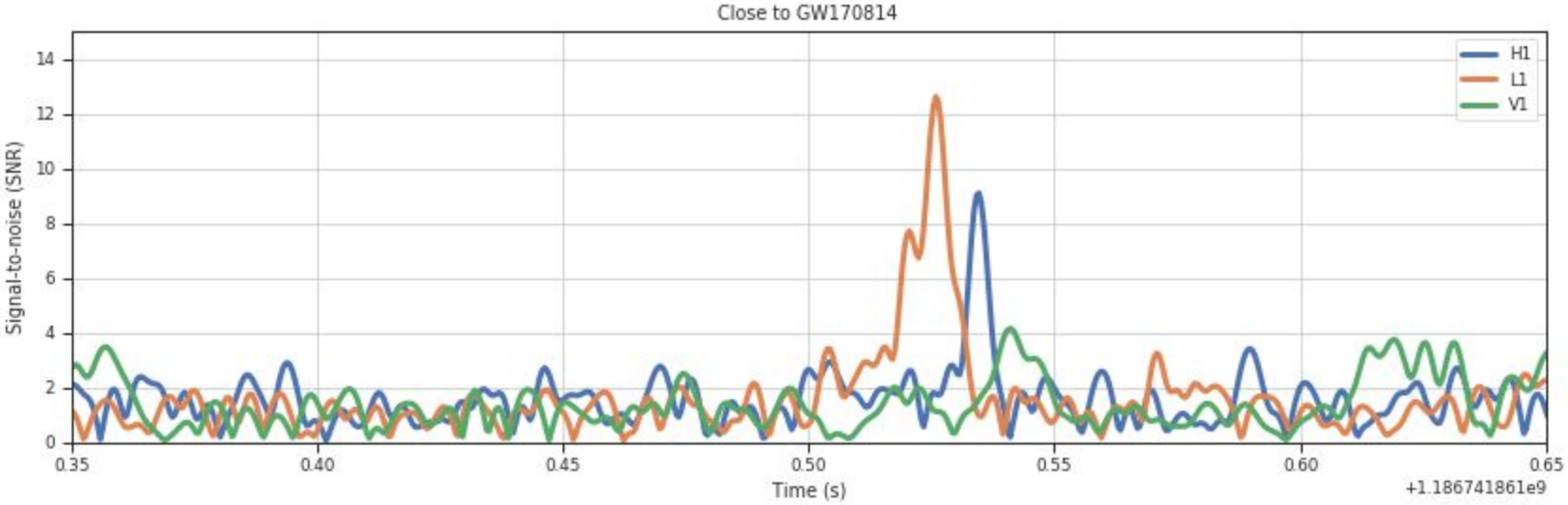
- For a glitch, or a signal that does not match the template, the expected value of the chi-test is increased by a factor proportional to the total SNR^2.
- For signals, the expected chi-square value is defined as follows:

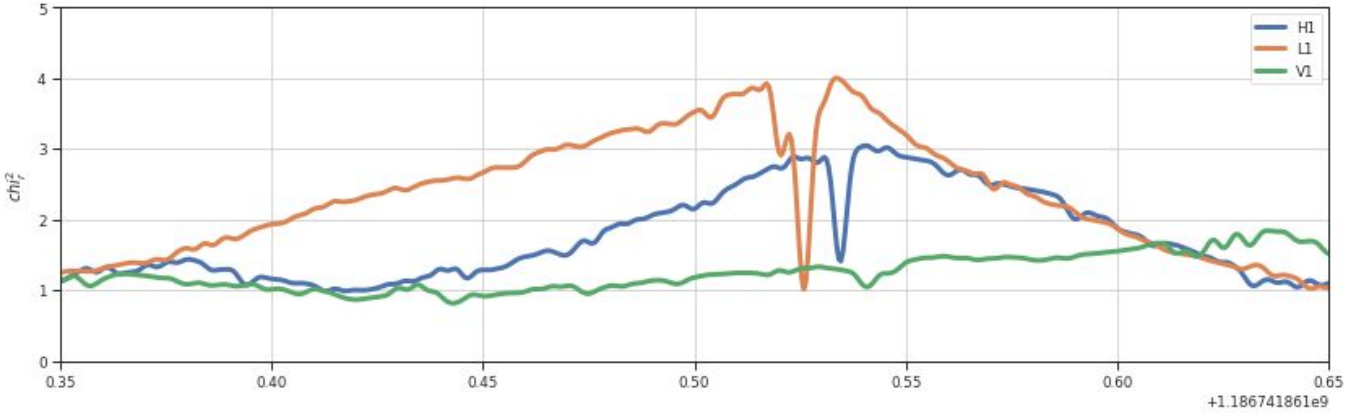
$$\langle \chi^2 \rangle = (2p - 2) + \epsilon^2 \rho^2$$

For low-mass CBC searches p = 16 has been chosen. Epsilon is the mis-match between signal and template.

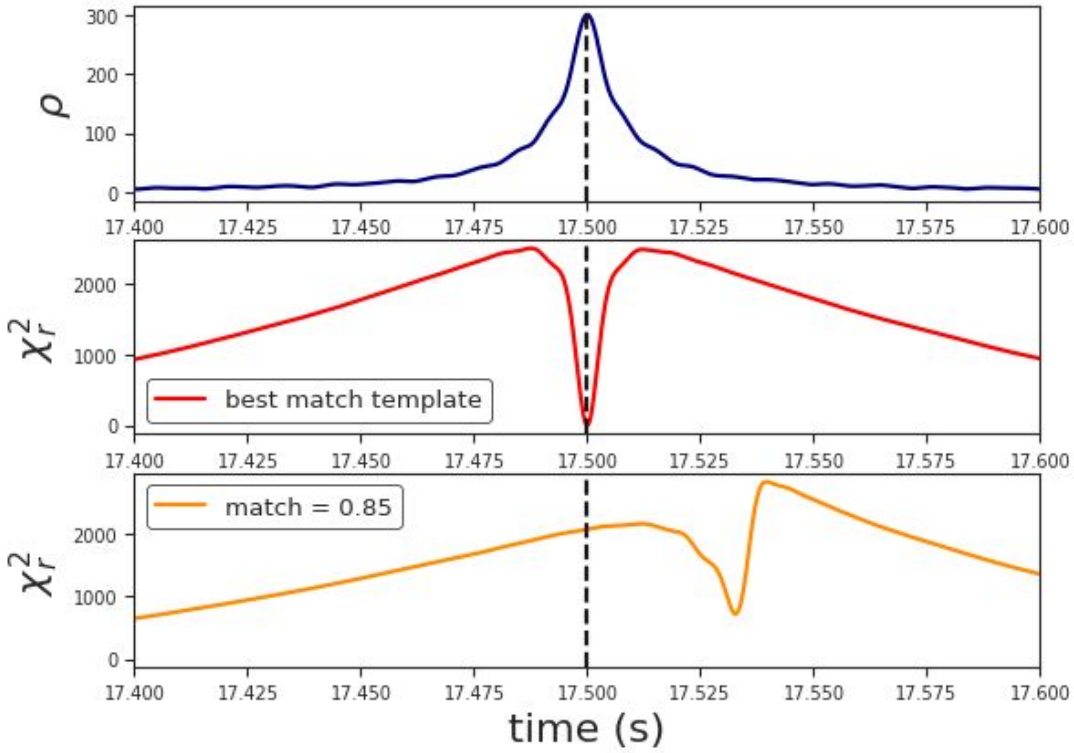


The figure shows the SNR and chi-square as a function of time. for a simulated CBC signal with SNR = 300 in a stretch of S5 data from the H1 detector. The SNR shows a characteristic rise and fall around the signal. **The chi-square is small at the time of the signal, but increases steeply to either side as the template waveform is offset from the signal in the data.**


Close to GW170814
Signal-to-noise (SNR)
Time (s)
+1.186741861e9
H1
L1
V1


chi_r^2
+1.186741861e9
H1
L1
V1

Signal consistency Test: Chi-square test



- For a GW signal that matches the template waveform exactly, the sum of squared residuals follows the chi-square distribution with 2p−2 degrees of freedom.

$$\langle \chi^2 \rangle = (2p-2) + \epsilon^2 \rho^2$$

- Waveform model = TaylorT4., lower cut-off frequency = 30.
- Injected signal: mass1 = 10, mass2 = 6.
- Nearest template: mass1 = 10.05, mass2 = 6.05

Signal based vetoes

How similar the SNR time-series of the data is to the SNR time series expected from a real signal ?

ξ^2 Test

$$Z_\alpha(t) = \rho_{2\alpha-1}(t) + i\rho_{2\alpha}(t) : \alpha \in [1, N_T]$$

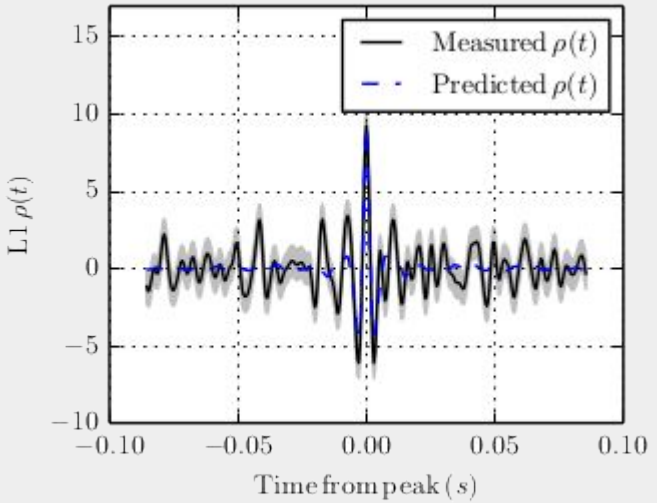
$$R_\alpha(t) = \frac{|h_{2\alpha-1}(f)|^2 + |h_{2\alpha}(f)|^2}{S_h(|f|)} e^{2\pi i f t}$$

$$\xi^2_\alpha(t) = |Z_\alpha(t) - Z_\alpha(0)\, R_\alpha(t)|^2$$

$$\langle \xi^2(t) \rangle = 2 - 2|R_\alpha(t)|^2$$

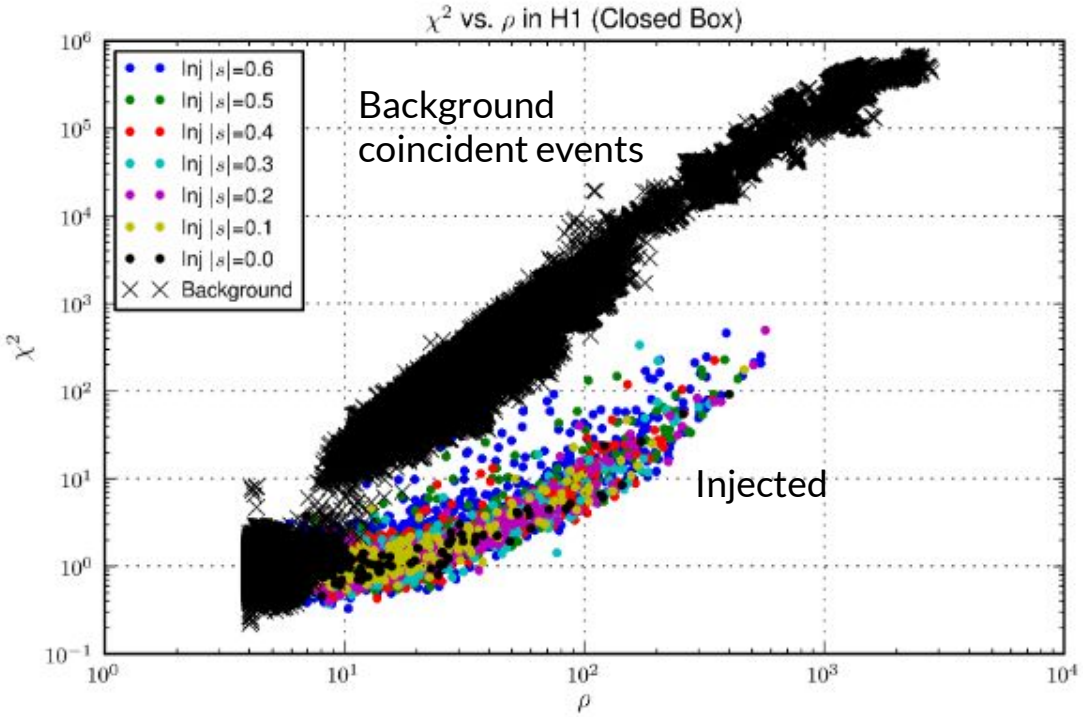
$$\xi^2_\alpha = \frac{\int_{-\delta t}^{\delta t} (\xi^2_\alpha(t)) dt}{\int_{-\delta t}^{\delta t} (\langle \xi^2(t) \rangle) dt}$$

Messick et al. (2017)



ξ^2 test is different from the traditional χ^2 test.

Interpretation of signal consistency Test



Credit : GstLAL group

ξ^2 vs SNR distribution for background and injections in Hanford.

It should be able to help separate background from injections in presence of non-Gaussian noise

GW data analysis

Stage-1: Detection of GW signals (Based on statistical measures)

Scientific investigations:

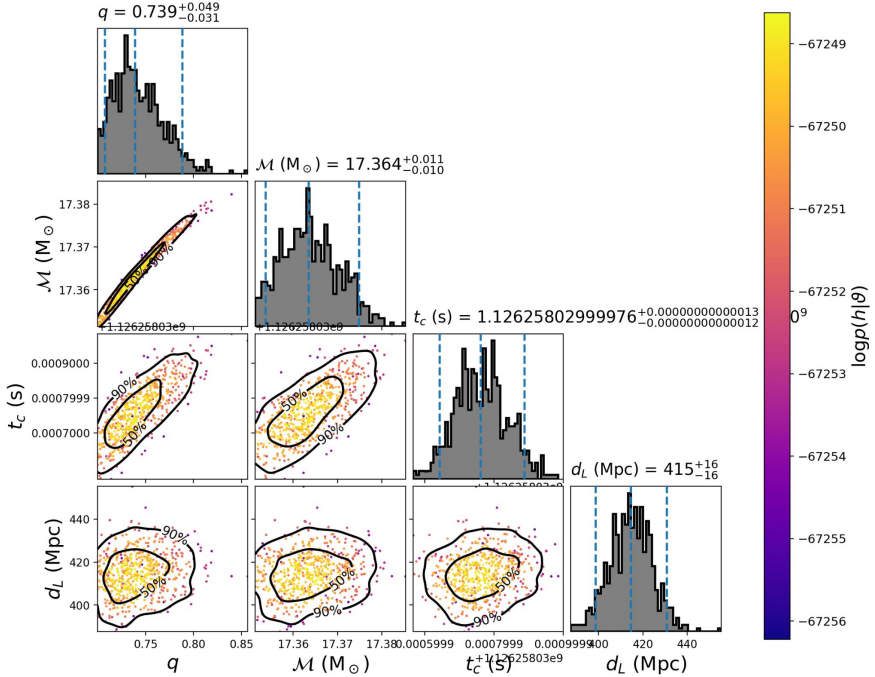
Level-1 :

- Suitable choice of template bank.
- Generation of the analytical waveforms.
- Matched filter using template waveforms

Level-2:

- Autocorrelation chi-square.
- Early trigger clustering.
- Event ranking.
- Mapping to FAR.

Stage 2: Parameter Estimation (Based on Bayesian Inference)



Parameter Estimation pipelines

GWTC-2 PE

Parameter	GW170104	GW170608	GW170729	GW170809	GW170814	GW170818	GW170823
$\mathcal{M}^{\mathrm{det}}\,(M_\odot)$	$25.2^{+1.7}_{-1.6}$	$8.5^{+0.06}_{-0.05}$	51.7^{+8}_{-9}	$29.9^{+2.2}_{-1.8}$	$27.2^{+1.2}_{-1.2}$	$32.2^{+2.8}_{-2.8}$	$39.1^{+4.7}_{-4.5}$
$m_1{}^{\mathrm{det}}\,(M_\odot)$	$37.3^{+8.2}_{-6.8}$	$12.0^{+6.0}_{-2.1}$	$74.5^{+13}_{-13.8}$	$41.9^{+10.3}_{-6.8}$	$33.9^{+6.3}_{-2.8}$	$43.5^{+9.7}_{-6.1}$	$52.7^{+12.7}_{-8.1}$
$m_2{}^{\mathrm{det}}\,(M_\odot)$	$22.9^{+5.9}_{-4.9}$	$8.0^{+1.6}_{-2.3}$	$48.8^{+14.6}_{-16}$	$28.7^{+5.9}_{-6.6}$	$28.9^{+2.6}_{-4.4}$	$32^{+5.9}_{-7.6}$	$39.1^{+7.8}_{-10.6}$
$\mathcal{M}^{\mathrm{src}}\,(M_\odot)$	$21.2^{+1.9}_{-1.4}$	$7.96^{+0.19}_{-0.19}$	$34.1^{+6.4}_{-4.5}$	$24.9^{+2.1}_{-1.5}$	$24.3^{+1.4}_{-1.2}$	$26.7^{+2.2}_{-1.9}$	$29^{+4.2}_{-3.2}$
$m_1{}^{\mathrm{src}}\,(M_\odot)$	$31.4^{+7.6}_{-6.0}$	$11.3^{+5.6}_{-2.0}$	$49.5^{+12.1}_{-10.2}$	$35^{+9.1}_{-5.9}$	$30.4^{+5.7}_{-2.7}$	$36.1^{+8.5}_{-5.3}$	$39.2^{+10.9}_{-6.6}$
$m_2{}^{\mathrm{src}}\,(M_\odot)$	$19.2^{+4.9}_{-4.0}$	$7.5^{+1.5}_{-2.2}$	$32.2^{+9.9}_{-9.1}$	$23.9^{+5}_{-5.3}$	$25.8^{+2.6}_{-4.0}$	$26.5^{+4.7}_{-6.0}$	$28.9^{+6.3}_{-7.2}$
q	$1.63^{+0.84}_{-0.56}$	$1.5^{+1.65}_{-0.46}$	$1.53^{+0.93}_{-0.48}$	$1.46^{+0.85}_{-0.42}$	$1.17^{+0.46}_{-0.15}$	$1.36^{+0.76}_{-0.33}$	$1.34^{+0.85}_{-0.31}$
χ_{eff}	$-0.08^{+0.16}_{-0.17}$	$0.057^{+0.19}_{-0.06}$	$-0.34^{+0.21}_{-0.27}$	$-0.06^{+0.18}_{-0.16}$	$-0.08^{+0.12}_{-0.12}$	$-0.08^{+0.2}_{-0.24}$	$-0.07^{+0.22}_{-0.21}$
a_1	$0.35^{+0.48}_{-0.31}$	$0.32^{+0.47}_{-0.29}$	$0.60^{+0.34}_{-0.51}$	$0.34^{+0.53}_{-0.31}$	$0.53^{+0.42}_{-0.48}$	$0.56^{+0.38}_{-0.5}$	$0.44^{+0.48}_{-0.4}$
a_2	$0.47^{+0.45}_{-0.42}$	$0.43^{+0.49}_{-0.39}$	$0.57^{+0.38}_{-0.5}$	$0.4^{+0.51}_{-0.37}$	$0.46^{+0.47}_{-0.42}$	$0.5^{+0.44}_{-0.45}$	$0.45^{+0.48}_{-0.41}$
$d_L\,(Mpc)$	970^{+400}_{-410}	318^{+128}_{-109}	2980^{+1410}_{-1400}	1020^{+310}_{-390}	584^{+130}_{-186}	1030^{+420}_{-350}	1920^{+870}_{-860}

- PyCBC inference
- Bilby
- Lal inference

Bayesian Analysis for Parameter Estimation:

Computation of the posterior PDF for the model parameters

posterior distribution — Likelihood — prior probability

$$p(\vec{\theta}_i|d, M_i) = \frac{L(d|\vec{\theta}_i, M_i)p(\vec{\theta}_i|M_i)}{p(d|M_i)}$$

d is observed data
M_i is the competing models for the data
θ_i is the parameters

Evidence can be computed as

$$p(d|M_i) = \int_{V_i} d\vec{\theta}_i L(d|\vec{\theta}_i, M_i)p(\vec{\theta}_i|M_i)$$

V_i is the parameter space volume in model M_i .

Need to compute the Likelihood function and evidence.

$$\log L(\vec{\theta}) \equiv \log p(d|\vec{\theta}) = -\frac{1}{2}\left\langle d - h(\vec{\theta}) \middle| d - h(\vec{\theta}) \right\rangle$$

$$\langle a|b\rangle = 4\ \mathrm{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}\,\mathrm{d}f.$$

Sampling techniques

MCMC

Estimate posterior by stochastically wandering in parameter space

Start with a random position(intelligent guess) in parameter space

Propose and accept a new point with a probability,

$$r_s = \min(1, \alpha)$$

$$\alpha = \frac{Q\left(\vec{\lambda} \mid \vec{\lambda}'\right) p\left(\vec{\lambda}' \mid d, H\right)}{Q\left(\vec{\lambda}' \mid \vec{\lambda}\right) p(\vec{\lambda} \mid d, H)}$$

Nested Sampling

Sample live points(say m) from prior.

Replace lowest L_* live point with a new point, found using MCMC starting from one of the remaining points, subjected to the constraint $L > L_*$.

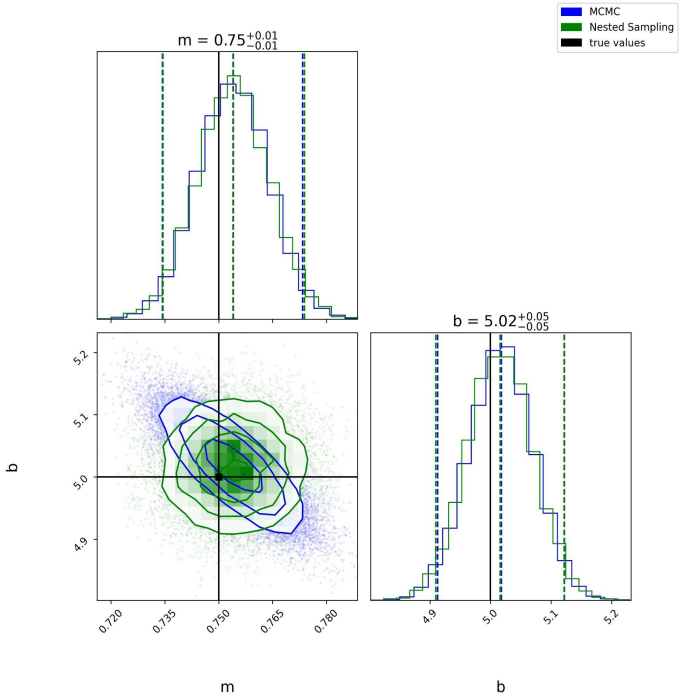
Repeat the 2nd step again until some stopping criterion is met.

Each discarded point is saved.

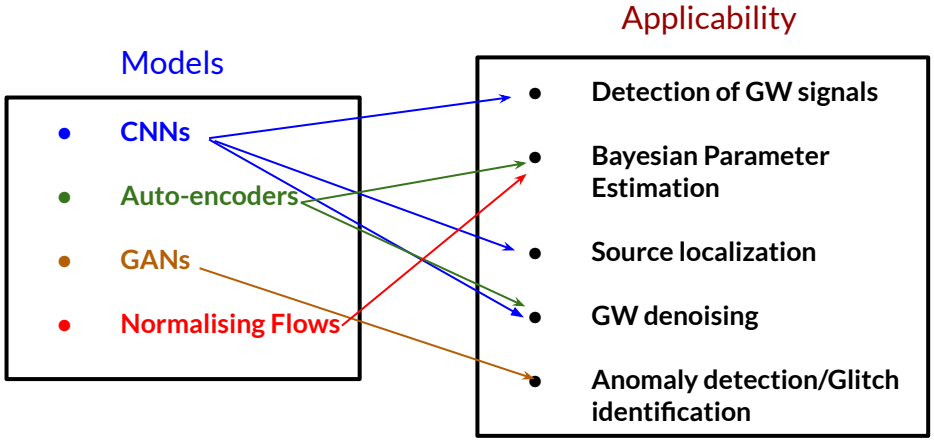
$$X(L_*) = \int_{p(d|\vec{\lambda},H)>L_*} d\vec{\lambda} p(\vec{\lambda} \mid H)$$

$$Z = \sum_{i=1}^{M} \frac{1}{2}(X_{i+1} - X_i) L_i$$

$$p(\vec{\lambda} \mid d, H) = \frac{\frac{1}{2}(X_{i-1} - X_{i+1}) L_i}{Z}$$



Current state of the art of application of ML



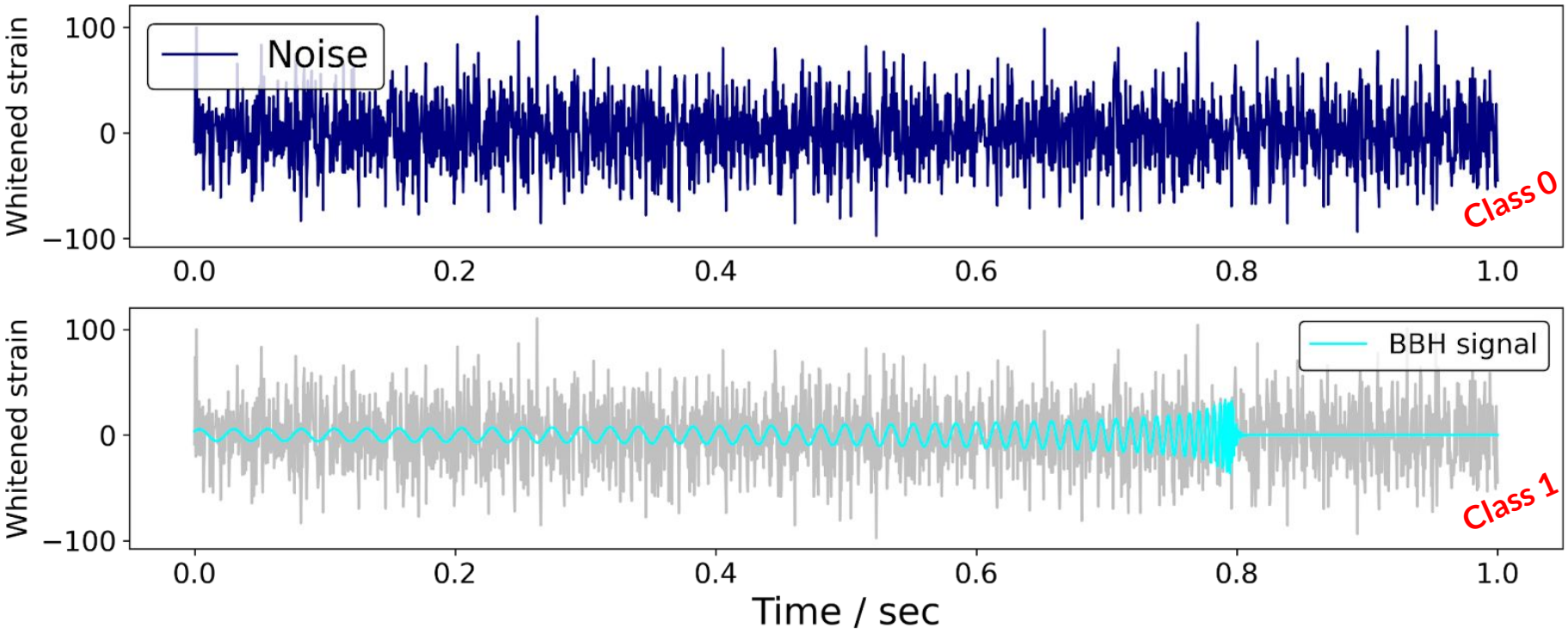
GW sources

- Compact Binary Inspiral Gravitational Waves
  a. BBH signal
  b. BNS signal
- Burst Gravitational Waves
- Continuous Gravitational Waves
- Stochastic Gravitational Waves [?]

Variation based on models / problems

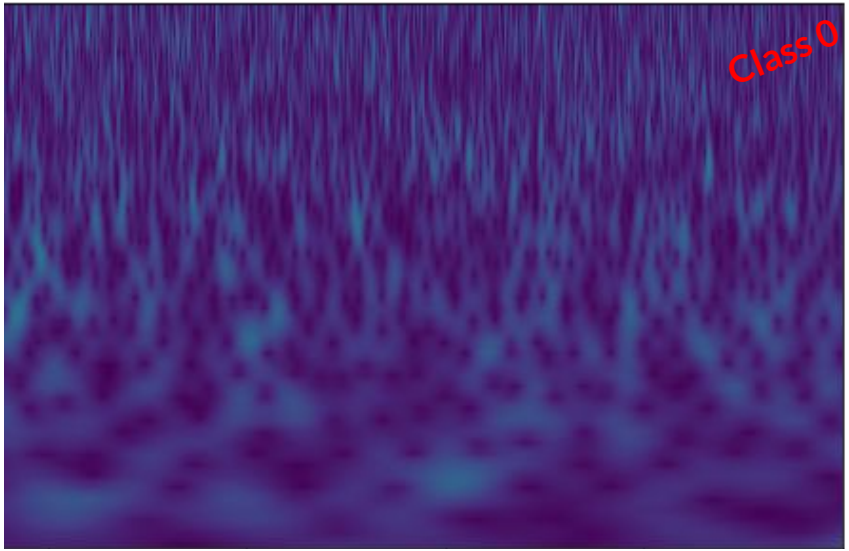
- Autoregressive neural network flows: Applied for the rapid likelihood free inference of BBH.
- Conditional variational autoencoders (CVAE) : Applied for the training on BBH signal to return Bayesian posterior probability estimates.
- Nested sampling with Normalising flows: It is included in the nested sampling's sampler setting to compute the likelihood quickly. Results are compared with DYNESTY. Good agreement between the computed log-evidences.
- Statistically informed deep learning: It is used for PE estimation of BBH signal. It stands for Constructive learning + Normalising flow + Wavenet Architecture. Results are shown for GW150914, GW170104, GW170814, GW190521.
- A resampling white box approach: A Morlet wavelet transformation is used to obtain a time-frequency image representation of a time-series strain data. Further CNN is used for analysis of those images.
- Deep transfer learning: It stands for Deep learning + transfer learning. It is used for the classification of GW signal from glitches. The transfer learning can be used to extract the features of the glitches and useful to find new classes of glitches in an unsupervised manner.
- Genetic algorithm optimized Neural Networks: It optimized the hyperparameter based on Genetic algorithm. Genetic algorithm can identify high quality architectures even when the initial hyperparameter seed values are far from a good solution.

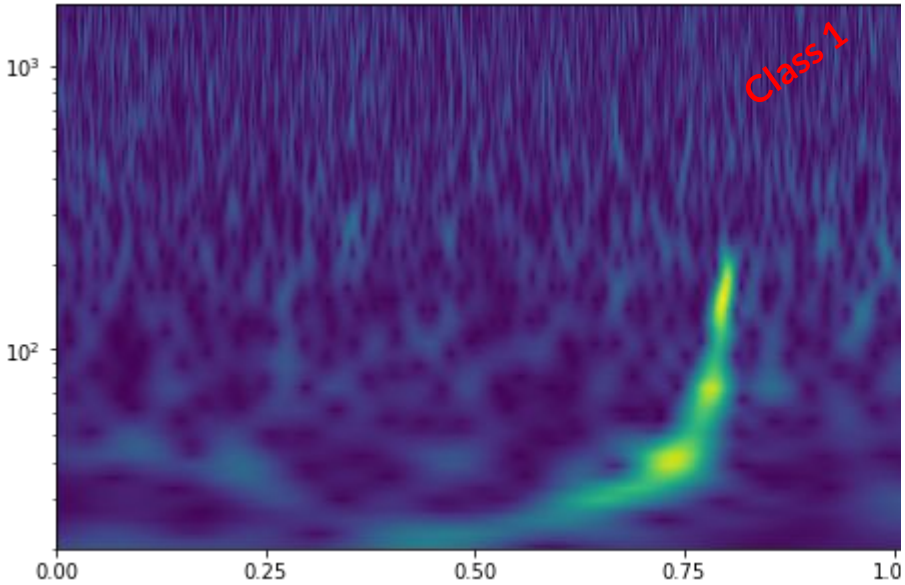
Detection of BBH signal using CNN:



Casting of a GW signal detection problem as a binary classification

Incorporating spectrogram images for training





CNN architecture for the search of BBH GW signal H. Gabbard et al. [PRL, 120,141103, 2018].

Motivation

- Matched Filtering is the standard method for the search of transient from BBH sources
- Computational Cost for low-latency search will increase with low cut-off frequency.

- CNN architecture can provides a fast signal identification scheme as the majority of calculations are performed prior to data taking during a training process.
- CNN can replace the matched filtering based search scheme for GW signal from BBH sources.

- The benefit of fast inference of CNNs—they analyze detector recordings much faster than real-time.
- Real-time alarms can provide useful hints for follow up searches of electromagnetic counterparts as well as for focused analysis with Bayesian parameter estimation.

Approach

- Only whitened time series of measured gravitational-wave strain as an input.
- Train and test on simulated binary black hole signals in synthetic Gaussian noise representative of Advanced LIGO sensitivity.
- It directly returns ranking statistics, equivalent to the inferred probability that data contains a signal.
- They showed that DNN network can classify signal from noise with equal search sensitivity as obtained from the matched filtering scheme.

Caution

- CNNs do not enjoy theoretical guarantees for stationary Gaussian data like matched filtering.
- They can, in principle, incorporate mechanisms to better deal with common non-Gaussianities in the data.

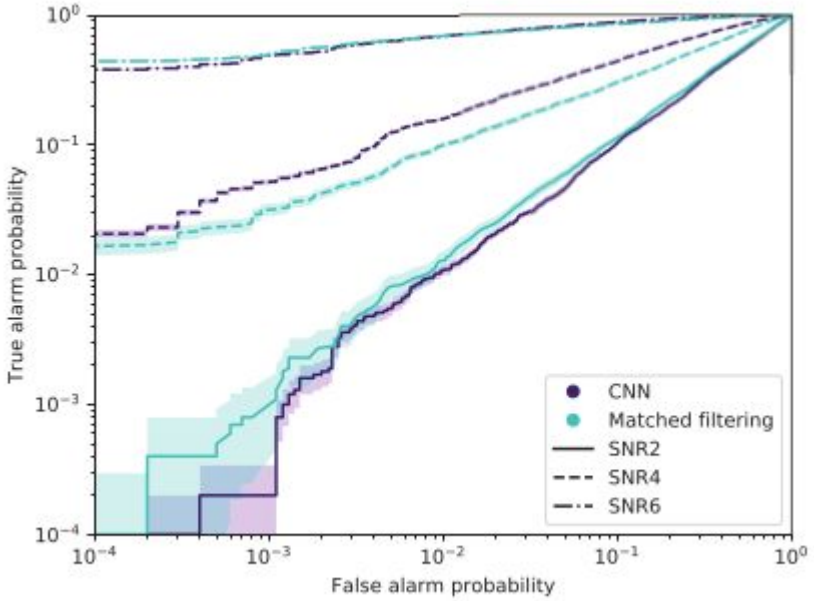
Simulation Details

- It consist of “whitened” simulated GW time series.
- IMRPhenomD type waveforms have been used for this study.
- Component black hole masses in the range from 5-95 solar mass has been chosen with zero spin.
- The simulated time series were chosen to be 1 sec in duration sampled at 8192 Hz.
- Training datasets contain 4×10^5 independent time series with 50% containing signal+noise and 50% noise-only.
- Each data sample time series is represented in the form of a 1×8192 pixel image with the gray-scale intensity.
- 90% of these samples for training, 5% for validation, and 5% for testing.

Parameter (Option)	Layer								
	1	2	3	4	5	6	7	8	9
Type	C	C	C	C	C	C	H	H	H
No. Neurons	8	8	16	16	32	32	64	64	2
Filter Size	64	32	32	16	16	16	n/a	n/a	n/a
MaxPool Size	n/a	8	n/a	6	n/a	4	n/a	n/a	n/a
Drop out	0	0	0	0	0	0	0.5	0.5	0
Act. Func.	Elu	Elu	Elu	Elu	Elu	Elu	Elu	Elu	SMax

- 6 convolutional layers, followed by 3 hidden layers is used in the final CN networks.
- Max-pooling is performed on the 1st, 5th, and 8th layer.
- Dropout is performed on the two hidden layers.
- Each layer uses an exponential linear unit (Elu) activation function.
- The last layer uses a Softmax activation function.

Performance: Matched Filtering Vs CNN



The figure described the ROC curves for test datasets containing signals with optimal SNR = 2, 4, 6. The true alarm probability versus the false alarm probability estimated from the output of the CNN (purple) and matched-filtering (cyan) approaches have been plotted [1].

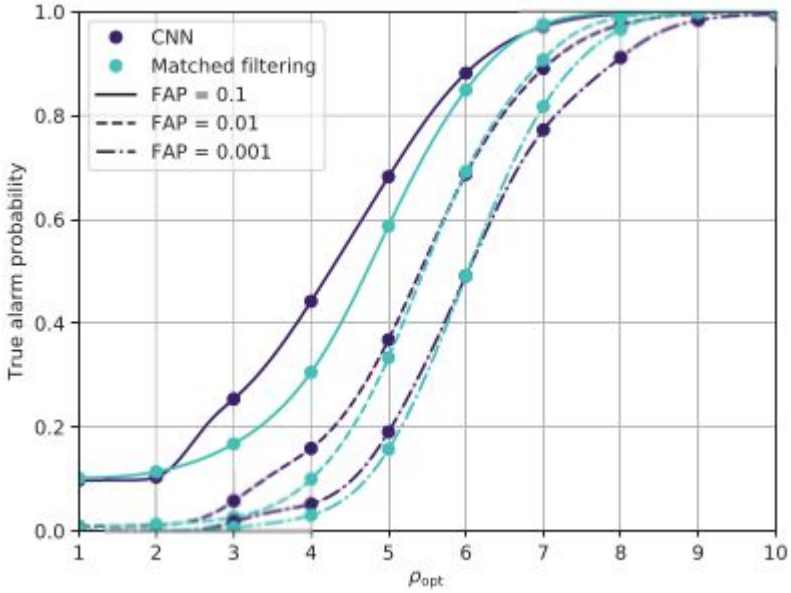
- False alarm probability: Fraction of noise incorrectly identified as signals.
- True alarm probability: Fraction of signal samples correctly identified.

- The results obtained from ROC curve showed that the CNN approach closely matches the sensitivity of matched-filtering for all test datasets across the range of false alarm probabilities explored in this analysis.

$$\text{True Positive Rate (TPR)} := \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{False Positive Rate (FPR)} := \frac{\text{FP}}{\text{FP} + \text{TN}}.$$

Performance: Matched Filtering Vs CNN



Efficiency curves comparing the performance of the CNN and matched-filtering approaches for different false alarm probabilities [1].

- The true alarm probability is plotted as a function of the optimal SNR for the CNN and the matched-filtering analyses.
- Solid dots indicate at which SNR values analyses were performed and line thicknesses are indicative of the statistical uncertainties in the curves.
- A good agreement between the approaches at all false alarm probabilities with the CNN sensitivity exceeding that of the matched-filter approach at low SNR and high false alarm probability has been observed.
- The matched-filter sensitivity marginally exceeds the CNN at high SNR and low false alarm probability. It can be resolved by increasing the number of training samples.

Conclusion

- This analysis represents a starting point from which a deep network can be trained on realistic non-Gaussian data. Since the claim of matched-filtering optimality is applicable only in the Gaussian noise case. Hence, there exists the potential for deep networks to exceed the sensitivity of existing matched-filtering approaches in real data.

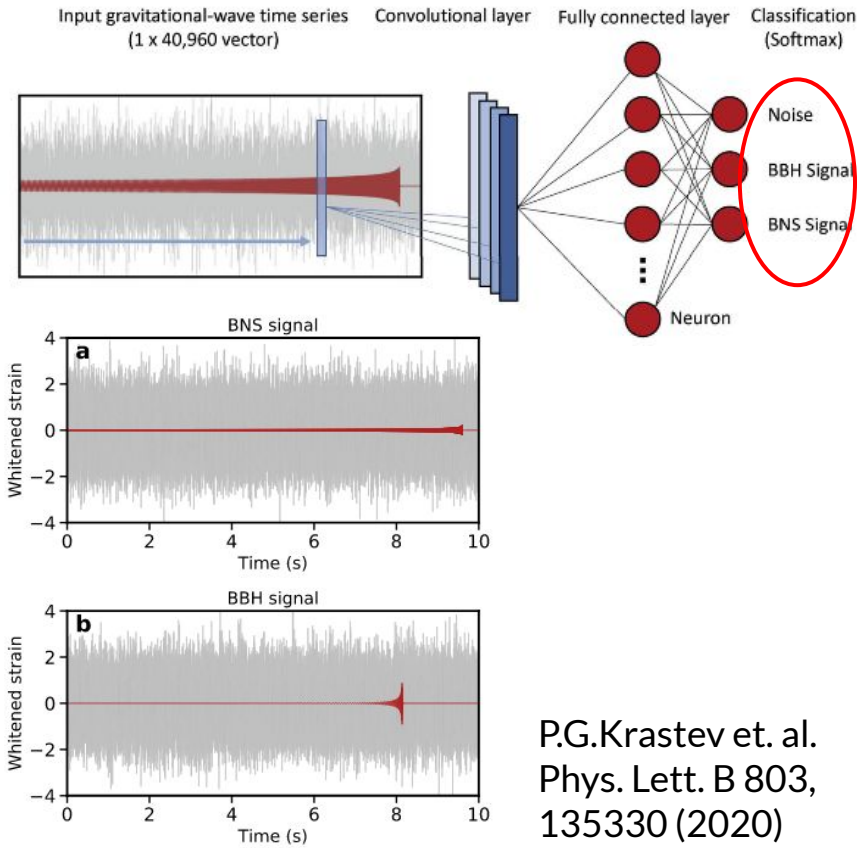
- This work presented results for BBH mergers, however, this method could be applied to other merger types, such as binary neutron star and neutron star-black hole signals.

- This supervised learning approach can also be extended to other well modelled gravitational-wave targets such as the continuous emission from rapidly rotating non-axisymmetric neutron stars .

- Finally there are possibilities for parameter estimation where in the simplest cases an output regression layer can return point estimates of parameter values.

Real-time detection of gravitational waves from binary neutron stars using artificial neural networks

Motivation



P.G.Krastev et. al. Phys. Lett. B 803, 135330 (2020)

Approach

- Only whitened time series of measured gravitational-wave strain as an input.
- The training sets used here consist of 100,000 independent time series with 1/3 containing BNS signal + noise, 1/3 BBH signal + noise, and 1/3 noise only.
- The validation and testing data sets each consist of 5,000 independent samples containing (approximately) equal fractions of each time-series class.
- To ensure that the neural network can identify BNS gravitational-wave signals over a broad range of astronomically motivated SNR values, they have started the network training with large SNR and then gradually reduce the SNR to lower levels.

Convolutional neural networks: A magic bullet for gravitational-wave detection?

What would be our level of confidence that there is a real event in the data when a binary classifier outputs a 1?

Issue with Binary Classification using CNN for GW searches:

1. Hard to identify the temporal location of potential signals in time series data of arbitrary length as generally we used a fixed length of data at the time of training.
2. The significance level obtained in the example based binary classification setup does not transfer easily to sliding-window based approaches for streaming data.
3. "false alarm rate" which can be derived from machine learning-based classifiers is directly linked to the training dataset.

Proposed Approach:

- They have extended the binary classification-based approach to longer inputs via a sliding window approach.
- The step size of the sliding window is also another parameter in the classifier problem

For example, assume the CNN outputs the sequence 1 – 1 – 0 – 1 – 1 – 0 – 1, where the coalescence happens roughly at the center value. How should these labels be counted as true (false) positives (negatives)?

-

T. D. Gebhard et al. [PRD, 100,063015, 2019].

Training and Testing datasets

For this work, they have created three datasets:

- A training dataset with 32768 examples,
- A validation set with 4096 examples,
- A testing dataset with 16384 examples.

This study is limited to waveforms from mergers of binary black holes, which are simulated using the effective-one-body model SEOBNRv4 in the time-domain.

Randomly sample values for the masses of the black holes, the z-components of their spins, the right ascension, declination, polarization, inclination, and coalescence phase angle, as well as the injection SNR have been used.

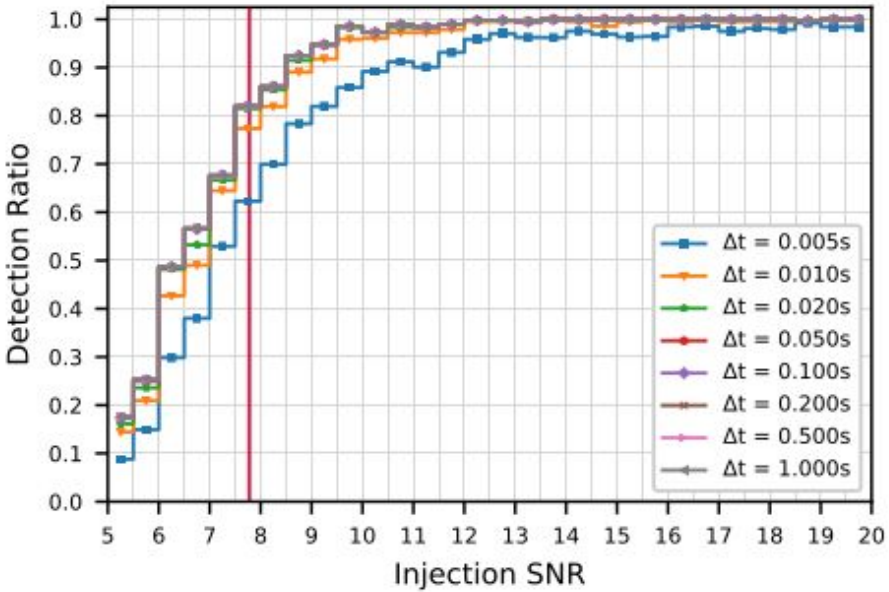
They have chosen a fully convolutional architecture. This means there are no fully connected (or dense) layers. Instead, the neural network only learns convolutional filters (or kernels), which make no assumptions about the size of their input data.

Model Architecture:

- An architecture should be chosen such that the receptive field is large enough to cover a substantial part of the signal.
- An architecture which incorporates the idea of the sliding window approach
- Instead of evaluating the network for each detector separately, they want to stack the recordings from all observatories and treat them as different channels of a single, multidimensional input.

In practice, they have used a stack of 12 (convolutional) blocks, each based on a dilated convolutional layer with 512 convolutional kernels of size 2.

Performance Evaluation-I



The figure shows that the detection ratio increases steeply with the injection SNR and achieves essentially 100% roughly at an SNR of 11. The vertical red line indicates the network SNR threshold above which the PYCBC search pipeline considers events for further analysis.

- When evaluated on full test set, our trained model is able to successfully recover approximately 88% of all injections, while on average producing a false positive about once every 12.5 minutes.
- For a more fine grained analysis, they have splitted the positive examples (i.e., the ones that do contain an injection) in the test set into 30 bins based on their respective injection SNR.
- Then compute the detection ratio independently for each of these bins to investigate how the sensitivity of the proposed method scales with the faintness of the signals.
- They have computed the global inverse false positive rate (i.e., averaged over all SNRs) as a function of Δt.

Improving significance of binary black hole mergers in Advanced LIGO data using deep learning :Confirmation of GW151216

Findings:

- This work is the first ML-based search that recovers all the binary black hole mergers in the first GW transients catalog (GWTC-1).
- This work also makes a clean detection of GW151216, which was not significant enough to be included in the catalogue.
- They have added a new coincident ranking statistic (MLStat) to a standard analysis that was used for GWTC-1.
- MLStat incorporates information from this ML classifier into the standard coincident search likelihood used by the conventional search.
- This leads to at least an order of magnitude improvement in the inverse false-alarm-rate (IFAR) for the previously "low significance" events GW151012, GW170729 and GW151216.

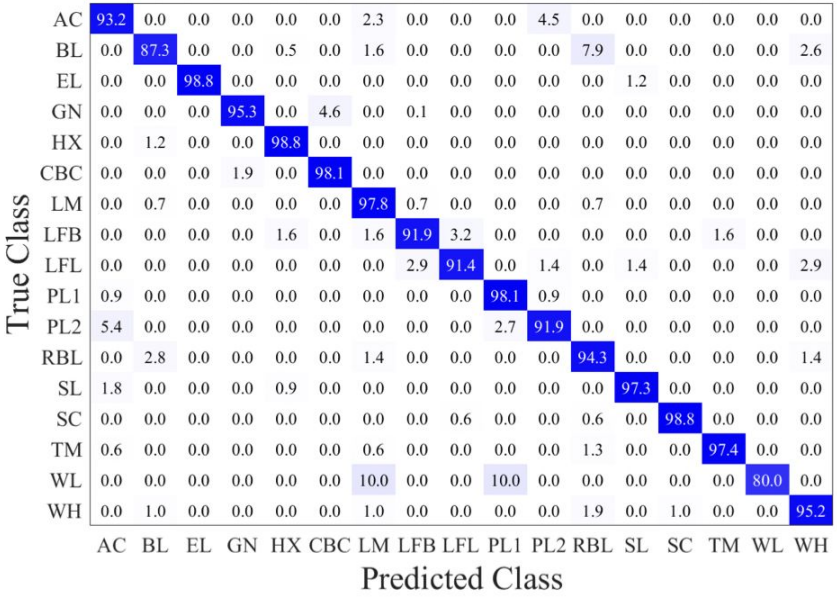
Proposed Approach:

- They have performed transfer learning to train "**InceptionV3**", a pre-trained deep neural network, along with curriculum learning to distinguish GW signals from noisy events by analysing their continuous wavelet transform (CWT) maps.
- The step size of the sliding window is also another parameter in the classifier problem

- The confidence in detection of GW151216 is further strengthened by performing its parameter estimation using SEOBNRV4HM_ROM.
- Considering the impressive ability of the statistic to distinguish signals from glitches, the list of marginal events from MLStat could be quite reliable for astrophysical population studies and further follow-up.

S Jadhav et. al. Phys. Rev. D 104, 064051 (2021)

Approach



Confusion matrix (in percent) for the combined validation data from all the levels of curriculum learning. A total of 17 transient classes have been used.

- They have created `CWT` scalograms by whitening and band-passing the data around a `GPS` trigger between 16 `Hz` and 512 `Hz`.
- They have considered a data slice of 1 second duration with the trigger time kept at centre, convert it into a scalogram and save it as a grayscale image with pixels denoting absolute values of `CWT` coefficients.
- As `Inception V3` is trained on natural images, the features extracted from different channels are most likely to differ from each other based on the biases in the images of natural objects.
- The choice of using grayscale colormap ensures the complete glitch morphology is saved in each of the three channels rather than getting divided based on the colormap.
- The network's convolution filters then observe the full evolution of a transient in each channel, and the channel-based biases are marginalised.

Construction of MLStat

The original likelihood ratio for detection of gravitational wave signal is given by:

$$\Lambda(H_\lambda|s) = \frac{p(s|H_\lambda)}{p(s|H_0)}$$

For each trigger, the proposed classifier gives the probability (P_{CBC}) of it belonging to the CBC class. They have updated the above likelihood ratio to include P_{CBC} as follows :

$$\Lambda'(H_\lambda|s) = \frac{p(s|H_\lambda) * \mathrm{P}_{\mathrm{H}}^{\mathrm{CBC}} * \mathrm{P}_{\mathrm{L}}^{\mathrm{CBC}}}{p(s|H_0)}.$$

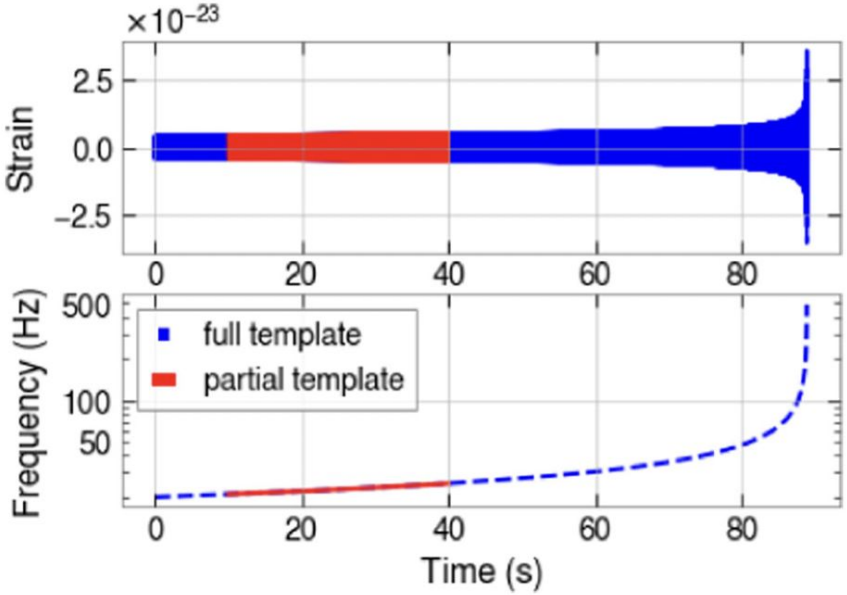
The new coincident ranking statistic (`MLStat`) can be written as follows:

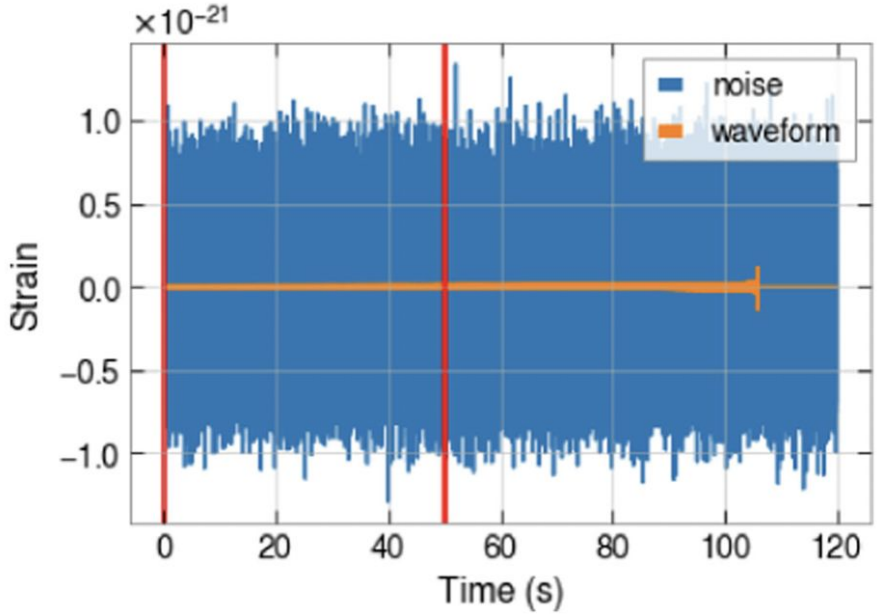
$$\tilde{\varrho}^2_{\mathrm{ml}} = \tilde{\rho}^2_c + 2 * \log(\mathrm{P}^{\mathrm{CBC}}_{\mathrm{combined}}),$$

- PyCBC workflow performs matched filtering on the data with a bank of templates.
- Triggers are then collected by thresholding and clustering the SNR time series.
- The SNR is re-weighed with two types of noise suppressing vetoes anda semi-coherent ranking statistic ensuring approximately trigger rate estimation across the search parameter space to determine the significance of events.
- They have analysed the triggers with their classifier and build the ML tool as an augmentation of the standard pipeline used by LIGO.

Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal

Objective: Detection of early inspiral GW signals from BNS system



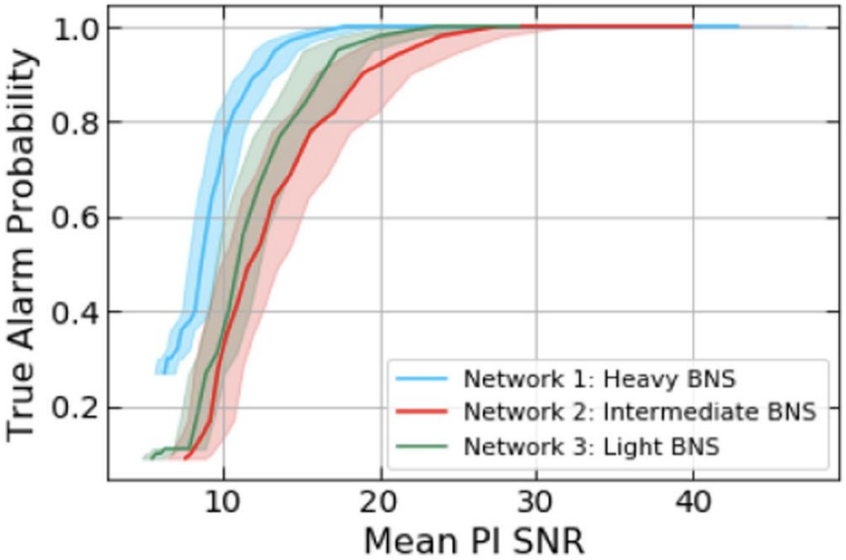


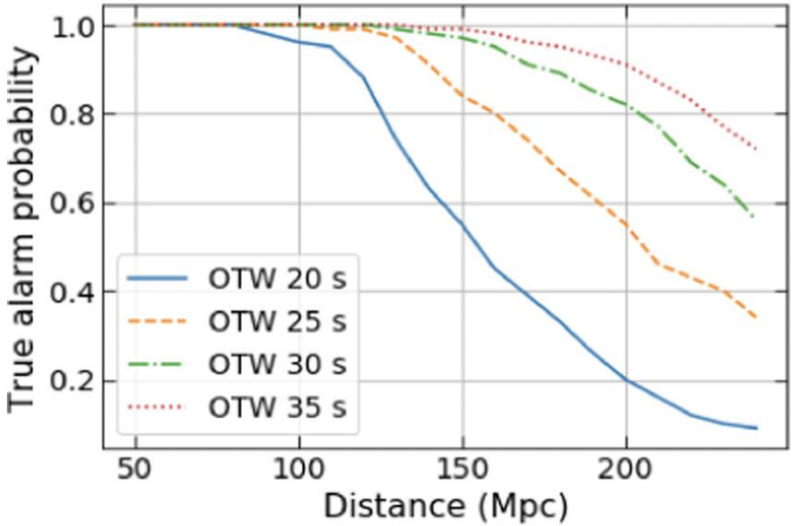
- Three different architectures are used.
- OTW is different for these architectures.

G.Baltus et. al. Phys. Rev. D 103, 102003 (2021)

Results

BNS	light	intermediate	heavy
$\mathcal{M}_c$ ($M_\odot$)	1.13 - 1.56	1.56 - 2.09	2.09 - 2.61
f_{low} (Hz)	20	20	20
Duration (s)	100 - 180	65 - 100	45 - 65
OTW (s)	80	50	30
Fraction of signal	0.44 - 0.8	0.5 - 0.77	0.46 - 0.66
Early alert before merger (s)	20 - 100	15 - 50	15 - 35





- Performance of the CNN trained on the heavy BNS systems for different OTWs.
- A longer window gives a higher number of detections.
- The detection accuracy can improve closer to the merger.

Using deep learning to localize gravitational wave sources

- Localization is not possible with a single detector.
- A pair of detectors localizes to a ring on the sky.
- The effective single-site timing accuracy is defined as

$$\sigma_t = \frac{1}{2\pi\rho\sigma_f}$$

- It is further possible to narrow down this range of possibilities by using relative amplitude and phase information from the detected signals.
- The localization can be further improved by using more detectors with large separation baselines between each individual interferometer.

C Chatterjee et al. [PRD, 100,103025, 2019].

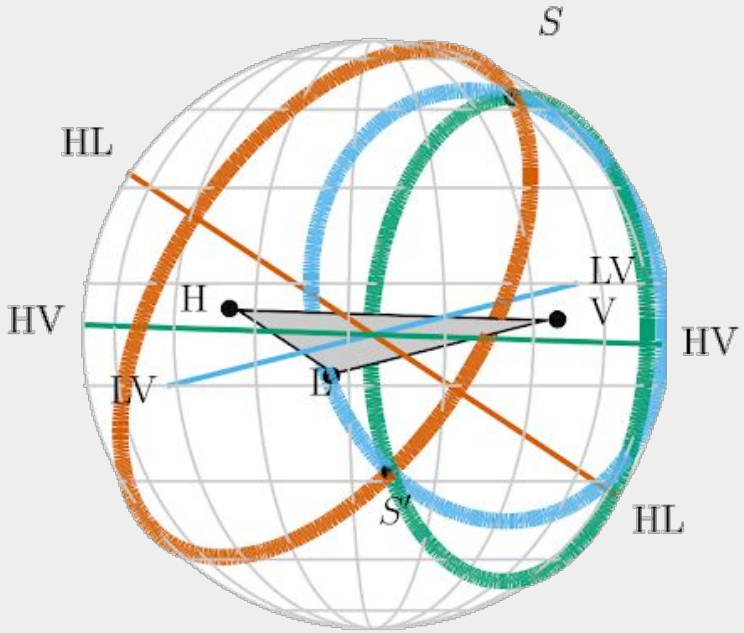
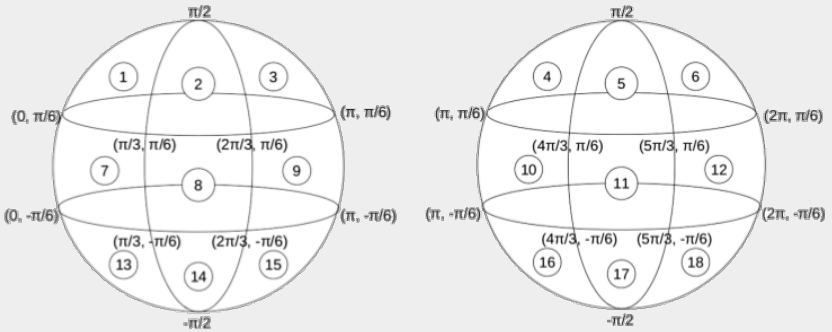


Image Credit: Fairhurst 2009

Proposed Approach

- They have modeled the sky as a sphere divided into many sectors.
- The sphere is divided into 18, 50, 128, 1024, 2048, and 4096 sectors.
- The sky direction of the gravitational wave source is estimated by classifying the signal into one of these sectors based on its right ascension and declination values for each of these cases.
- They have Injected simulated BBH gravitational signals of component masses sampled uniformly between 30 - 80 Solar Mass.
- SEOBNRv4 Waveform are used in this simulation.
- They have considered it as a **Multi-label Classification problem**.



Labeling convention for 18 sectors. The RA and Dec ranges for each of the labeled sectors are indicated.

Inclusion of extra parameters as input features

- Cross-correlation Definition:

$$w(t) = u(t) \otimes v(t) \triangleq \int_{-\infty}^{+\infty} u^*(\tau) v(\tau + t) d\tau$$

- The time corresponding to the maximum cross-correlation values gives the arrival time delay of one signal relative to the other.
- Analytic representations of TD signals obtained through a Hilbert transformation (HT).
- HT allows us to extract the instantaneous amplitudes, instantaneous phases, and phase lags between two signals, from noisy signals in the time domain.

Arrival time delays: H-L, L-V, H-V

Maximum Cross-correlations: HL, LV, HV

Average phase lags: H-L, L-V, H-V

Ratios of average amplitudes around merger: H/L, L/V, H/V

- They first find the instantaneous amplitudes around the merger region by searching for ten highest peaks from each GW sample for the three detectors and record the instantaneous phases corresponding to those peaks.
- They calculated the average of the ten instantaneous phases. This gives the average instantaneous phase around the merger of each signal.
- They calculated the differences of these instantaneous phases for each pair of detectors to get the instantaneous phase lags.

Experiments

They have generated 12,000 samples for pure GW signals without noise.

They have splitted it into a training and test set consisting of 10,800 samples and 1200 samples, respectively.

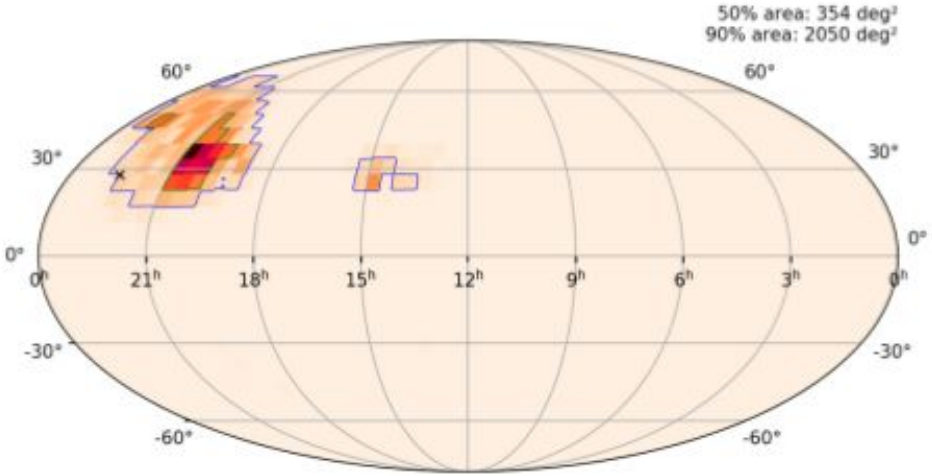
For signals plus Gaussian noise, we have trained the network with 160,000 samples and used a validation set of 40,000 samples and a test set of 4000 samples.

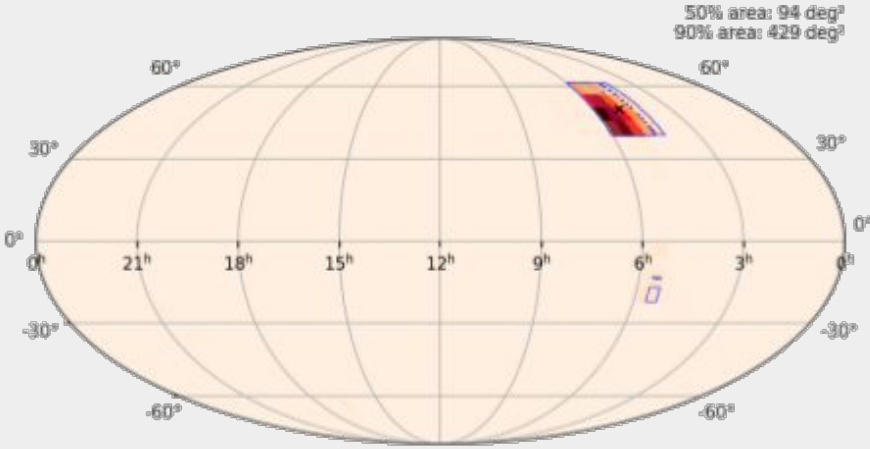
They have considered three different SNR ranges in experiments with signal plus Gaussian noise: [50–55], [20–35] and [10–110]. The third case was studied by applying curriculum learning.

SNR	Number of sectors	Training accuracy	Test accuracy	Revised test accuracy[a]
[50–55]	18	89%	91%	98.5%
[50–55]	50	80%	84%	98.25%
[50–55]	128	70%	77%	97.8%
[20–35]	18	80%	85%	97.27%
[20–35]	50	69%	73%	96%
[20–35]	128	55%	62%	92%
[10–110][b]	128	60%	65%	94.5%[c]

[a]Correct within one sector.
[b]With curriculum learning.
[c]Tested on samples with SNR of [20–35].

Test result for detected GW events





Probability heat map of localization with contours by ANN for GW170818, GW170823. The blue line shows the 90% contour, and the green line shows the 50% contour. The exact sky direction of the GW signal is marked with a black cross.

Conclusion

- They have built an ANN and have approached it like a classification problem, in which the sky, modeled as a sphere, has been divided into several sectors, and they have trained ANN model so that it classifies simulated GW signals into their correct sector based on its RA and Decs.

- They have reported more that 90% accuracy for all the cases they have considered. They tested their model's localization accuracy by using test samples with injection parameters of **GW150914**, **GW170818**, and **GW170823** and conclude that the model gives feasible results for Gaussian noise and advanced LIGO PSD.

- The major advantage of their approach is that it is **orders of magnitude faster than** any current localization technique.

- They also plan to work with higher numbers of sectors and ensure that the model has enough training examples for higher sector numbers by having a larger and more uniform training dataset.

- They will explore the performance of our network for binary neutron star mergers and lower mass black holes in the future.

List of Current Projects @ Nikhef / Utrecht on ML

- Anomaly Detection [Stefano, Melissa]
    a. GANs for Glitches. [Melissa, Vincent, Stefano][arxiv 2022]
    b. Classification and modeling of glitches. [Melissa, Jade]

- Early Warning for CBC signal [Gregory, Justin, Melissa , Harsh in collaboration with Prof. J. R. Cudell]
    a. Efficiency for single detector case (Time-domain) [Physics Rev. D. 103, 102003, 2021]
    b. Going down to frequency domain [arxiv 2022]

- Search for core-collapse supernova [Melissa in collaboration with GW group from Rome, Italy and Spain.]
    a. Deep learning for core-collapse supernova detection [PRD, 103,063011, 2021].

- Fast computation of waveforms from BBH system. [Stefano in collaboration with Prof. Walter Del Pozzo.]
    a. ML generator for Non-spin / Aligned-spin waveform. [PRD, 103,043020, 2021].

References:

1. Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders., D. George et al. [arxiv, 2017].
2. Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data, D. George et al. [Elsevier, 2017].
3. Classification and unsupervised clustering of LIGO data with Deep Transfer Learning, D. George et al. [PRD, 101501(R), 2018].
4. Matching matched filter with deep networks for gravitational wave astronomy, H. Gabbard et al. [PRL, 120,141103, 2018].
5. Convolutional Neural Network: A magic bullet for gravitational wave detection. T. D. Gebhard et al. [PRD, 100,063015, 2019].
6. Using deep learning to **localize gravitational wave sources**, C Chatterjee et al. [PRD, 100,103025, 2019].
7. Deep-learning **continuous gravitational waves.**, C. Dreissigacker, et al. [PRD, 100, 044009, 2019].
8. Generalized application of the Viterbi algorithm to searches for **continuous gravitational-wave signals**, J. Bayley et al. [PRD, 100,023006, 2019].
9. **Binary neutron stars** gravitational wave detection based on wavelet packet analysis and convolutional neural networks, B.J. Lin et al. [Springer 2019].
10. **Bayesian Parameter Estimation** using conditional variational autoencoders for gravitational wave astronomy, H Gabbard et al. [arxiv 2020].
11. Robust machine learning algorithm to search for **continuous gravitational waves**, J. Bayley et al. [PRD, 102,083024,2020].
12. Deep learning for gravitational-wave data analysis: A resampling white-box approach, M.D. Morales et al. [arxiv, 2020].
13. Exploring gravitational-wave **detection and parameter inference** using Deep Learning methods, J.D. Alvares et al. [arxiv, 2020].
14. Genetic-algorithm-optimized neural networks for gravitational wave classification, D.S. Deighan et al. [arxiv, 2020].

References:

1. Detection and classification of **supernova gravitational wave** signals: A deep learning approach, M. L. Chan et al. [PRD, 102,043022,2020].
2. Gravitational-wave **parameter estimation** with autoregressive neural network flows, S.R.Green et al. [PRD 102,104057, 2020].
3. Learning **Bayesian Posteriors** with Neural Networks for Gravitational-Wave Inference, A. J. K. Chua et al, [PRL, 124,041102, 2020].
4. **Nested Sampling** with Normalising Flows for Gravitational-Wave Inference, M. J. Williams et al. [arxiv 2021].
5. Statistically-informed deep learning for gravitational wave **parameter estimation, H Shen et al.** [arxiv 2021].
6. **Detection and Parameter Estimation** of Gravitational Waves from Binary Neutron-Star Mergers in Real LIGO Data using Deep Learning, P. G. Krastev et al. [arxiv, 2021].
7. Deep learning ensemble for real-time gravitational wave detection of **spinning binary black hole mergers**, Wei Wei et al. [PLB, Elsevier, 2021].
8. Deep Generative Models of Gravitational Waveforms via Conditional Variational Autoencoder, C. H. Liao et al.[arxiv, 2021].
9. Generalised gravitational burst generation with Generative Adversarial Networks, J. Mcginn et al. [arxiv, 2021].
10. Generalized Approach to Matched Filtering using Neural Networks, J Yan et al. [Phys. Rev. D **105**, 043006].

[Short summary: In this paper, the authors defined a relationship between the deep learning technique and the traditional matched filtering technique. They have showed a NN can be constructed analytically to exactly implement matched filtering, and can be further trained on data or boosted with additional complexity for improved performance.]