Machine Learning applications in Gravitational Wave research to classify transient signals

CERN seminars
(February 13th, 2019)

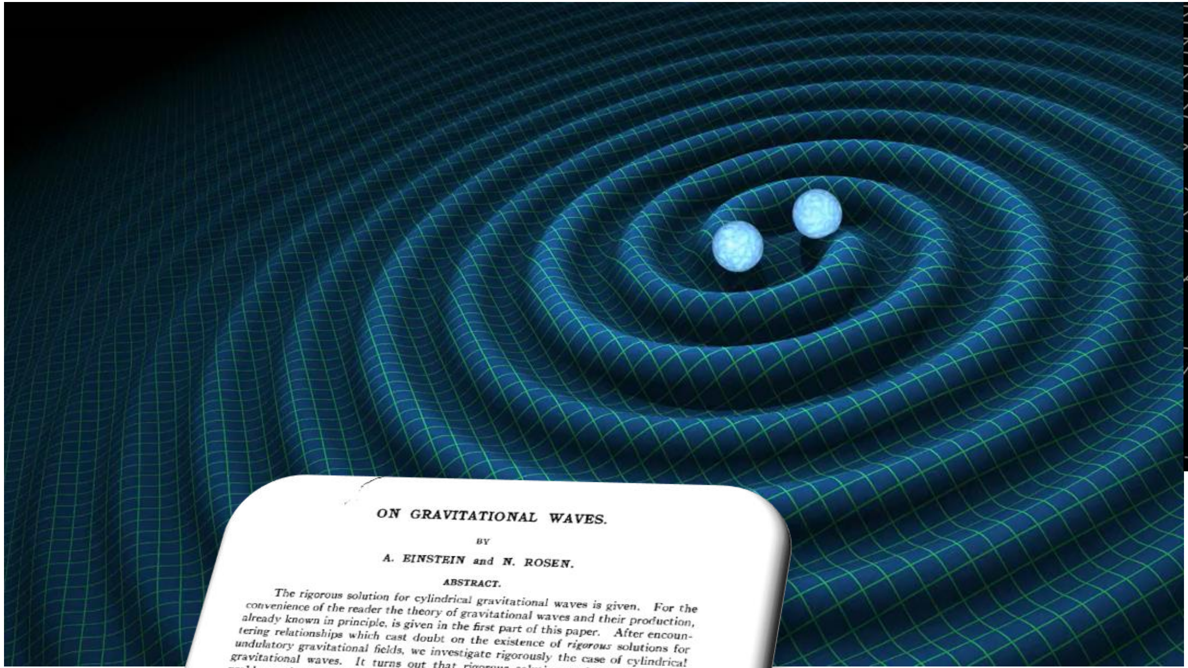
Elena Cuoco, EGO and SNS
www.elenacuoco.com
Twitter: @elenacuoco

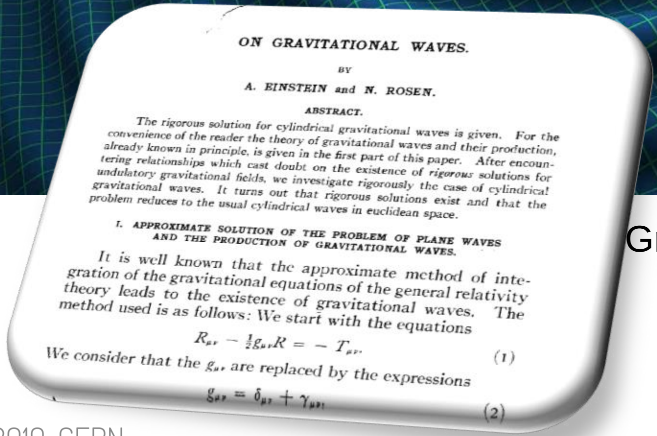
EUROPEAN GRAVITATIONAL OBSERVATORY

SCUOLA NORMALE SUPERIORE

VIR-0166A-19

What are Gravitational Waves (GWs)?




ON GRAVITATIONAL WAVES.

BY

A. EINSTEIN and N. ROSEN.

ABSTRACT.

The rigorous solution for cylindrical gravitational waves is given. For the convenience of the reader the theory of gravitational waves and their production, already known in principle, is given in the first part of this paper. After encountering relationships which cast doubt on the existence of *rigorous* solutions for undulatory gravitational fields, we investigate rigorously the case of cylindrical gravitational waves. It turns out that rigorous solutions exist and that the problem reduces to the usual cylindrical waves in euclidean space.

I. APPROXIMATE SOLUTION OF THE PROBLEM OF PLANE WAVES AND THE PRODUCTION OF GRAVITATIONAL WAVES.

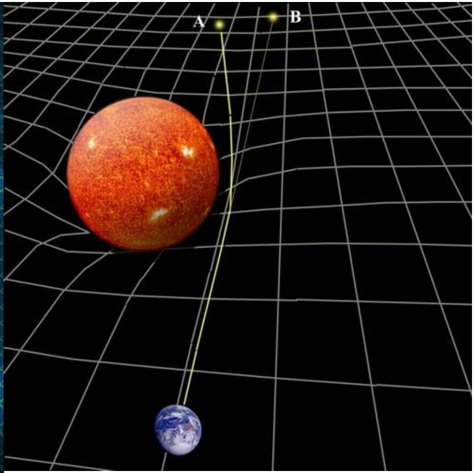
It is well known that the approximate method of integration of the gravitational equations of the general relativity theory leads to the existence of gravitational waves. The method used is as follows: We start with the equations

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -T_{\mu\nu}. \quad (1)$$

We consider that the $g_{\mu\nu}$ are replaced by the expressions

$$g_{\mu\nu} = \delta_{\mu\nu} + \gamma_{\mu\nu}, \quad (2)$$

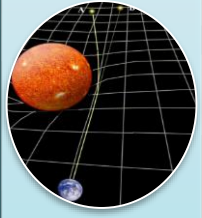
Gravitational Waves (1916)



General Relativity (1915)

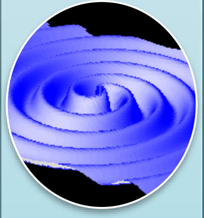
$$\mathbf{G}_{m} = \frac{8pG}{c^4}\mathbf{T}_{m}$$

A long history…



1915

- General Relativity

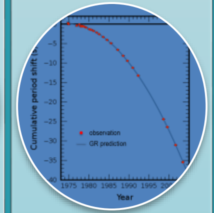


1916

- Gravitational Waves

1966

- Weber and Resonant Bars

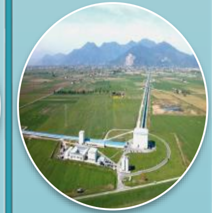


1974

- Hulse-Taylor:Observing the pulsar binary PSR B1913+16

1980-1990

- Cryogenic Resonant Bars

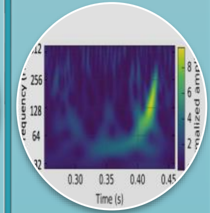


1993

- The approval of Virgo Experiment

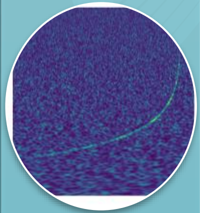
1999+

- Data taking from LIGO an Virgo



2015

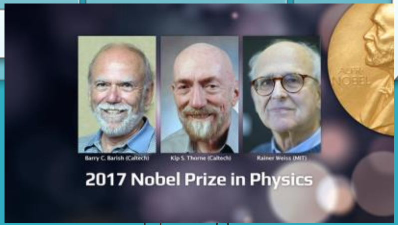
- GW BHBH detection



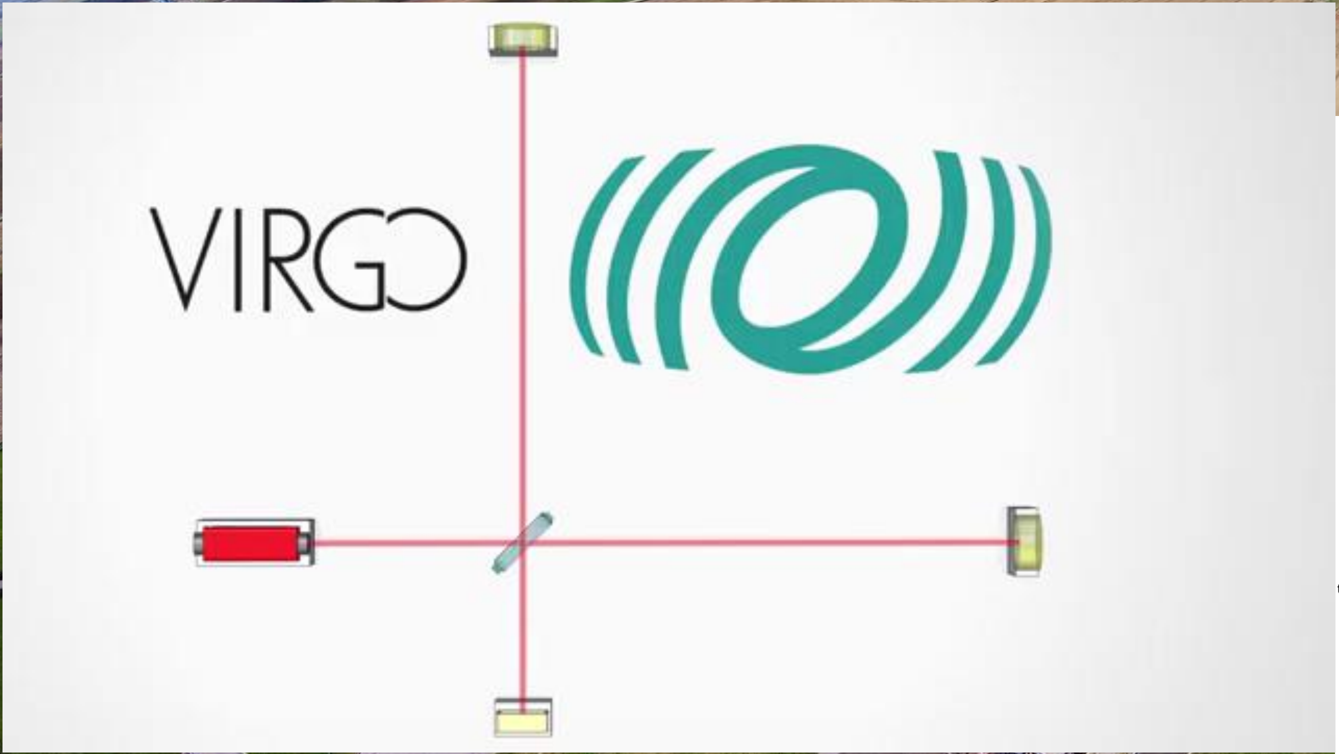
2017

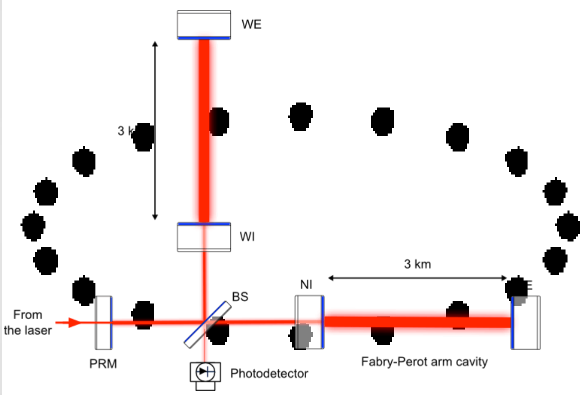
- BNS detection: Multi-messenger Astronomy

~100 years



How we detected GWs?





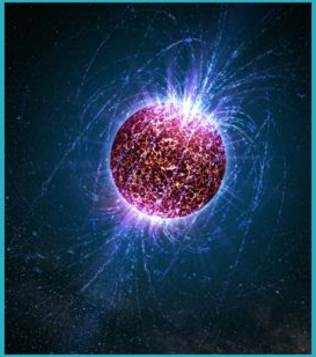
Astrophysical sources

Short➔ long

Known ➔ unknown form

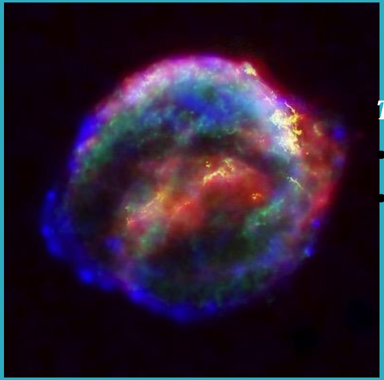
Coalescing Binary Systems CBC

- ✓ Black hole – black hole
- ✓ Neutron star – neutron star
- BH-NS
- Analytical waveform



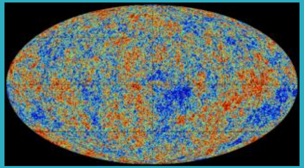
Continuous Sources

- Spinning neutron stars
- monotone waveform



Transient 'Burst' Sources

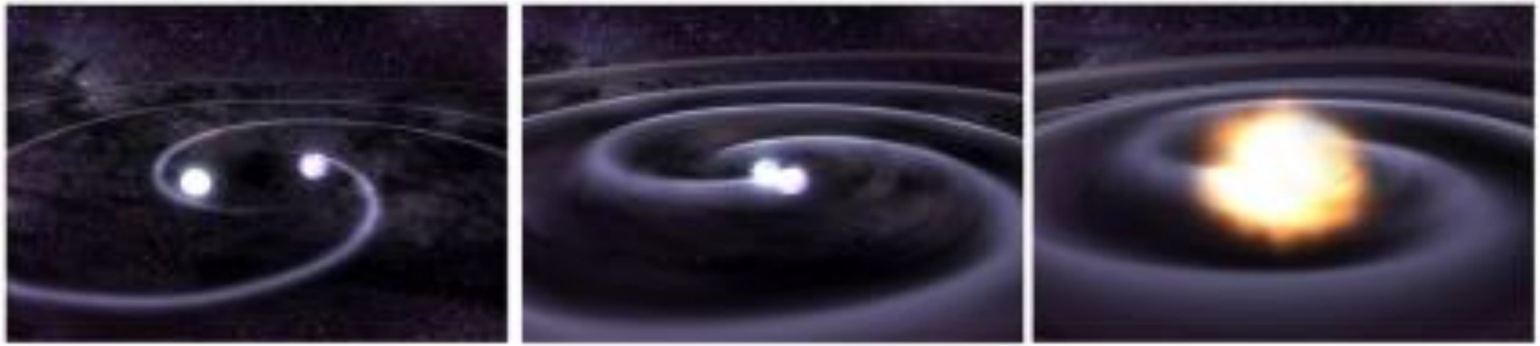
- core collapse supernovae
- unmodeled waveform



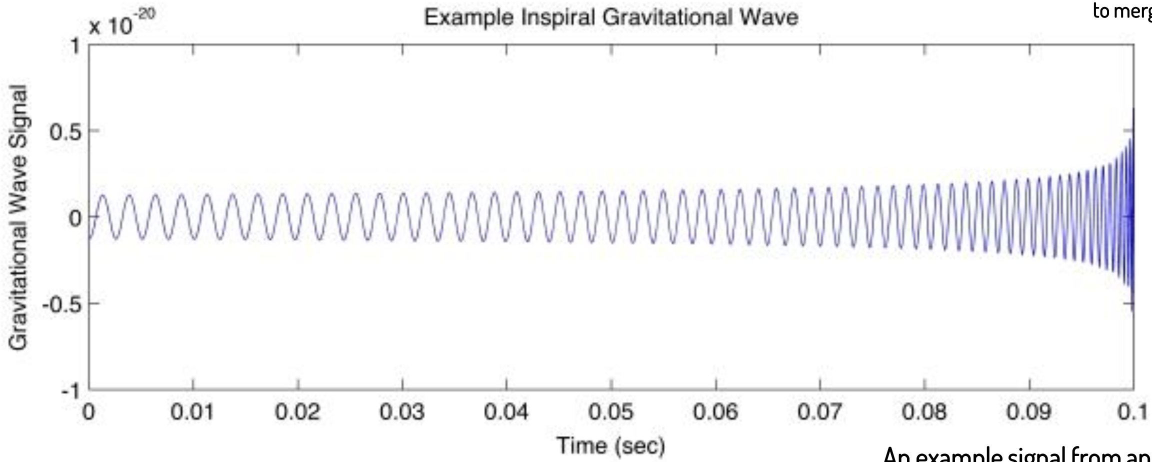
Cosmic GW Background

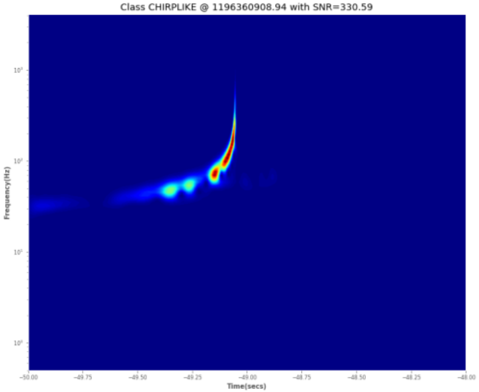
- residue of the Big Bang,
- stochastic, incoherent background

CBC Gravitational Wave signals



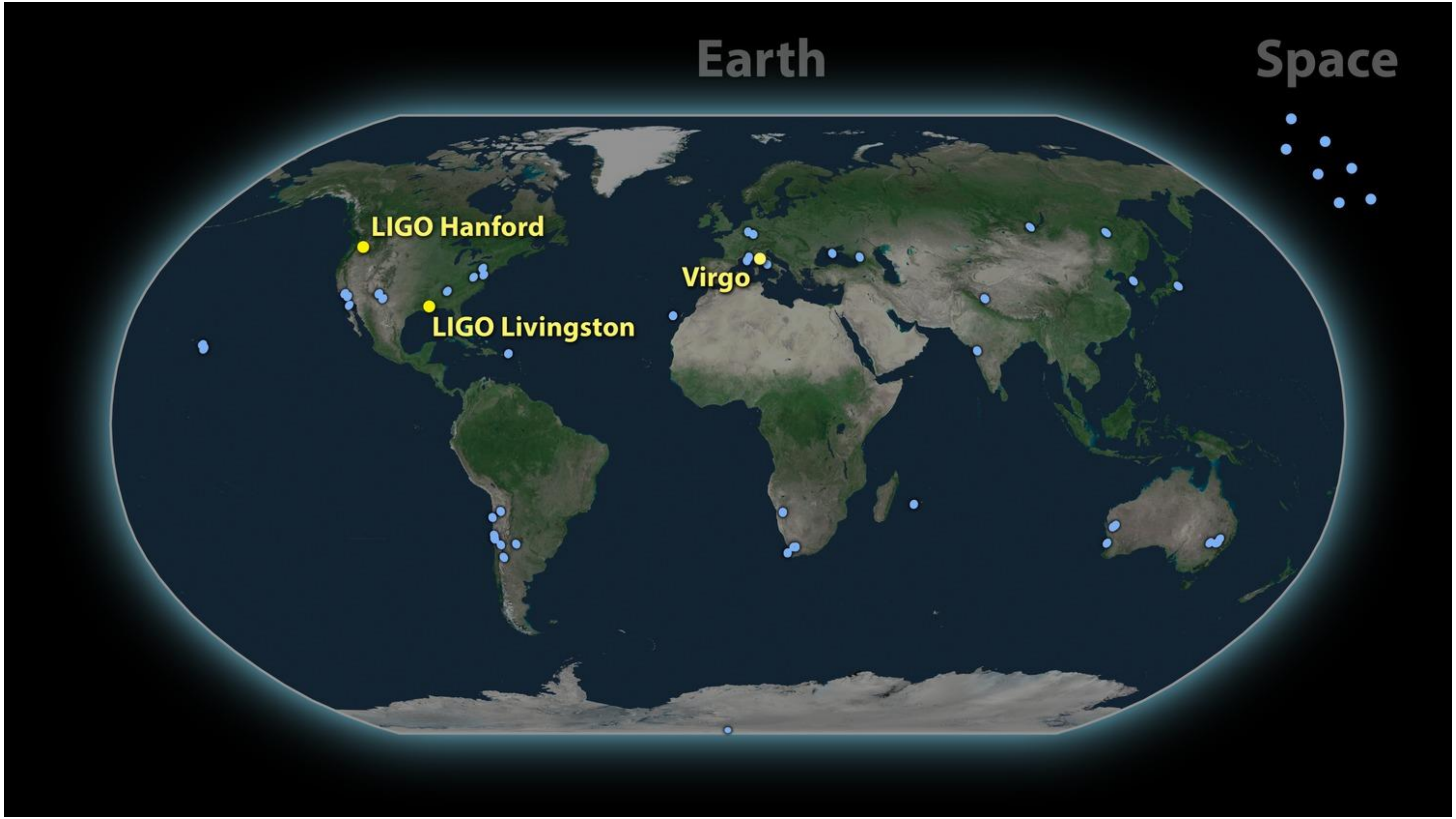
An artist's impression of two stars orbiting each other and progressing (from left to right) to merger with resulting gravitational waves. [Image: NASA/CXC/GSFC/T.Strohmayer]



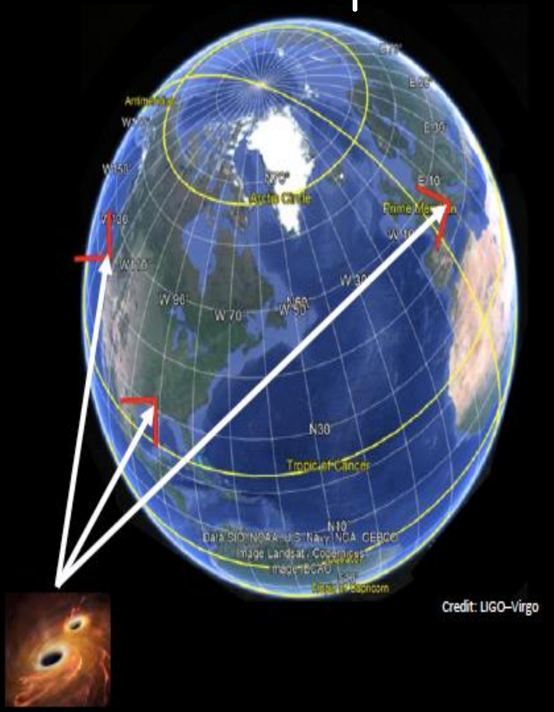


An example signal from an inspiral gravitational wave source. [Image: A. Stuver/LIGO]

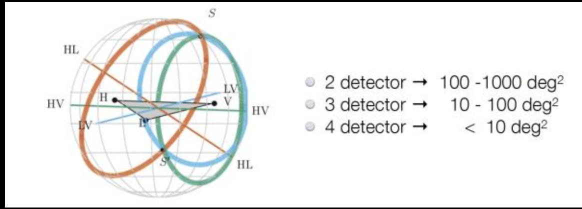
International Collaboration



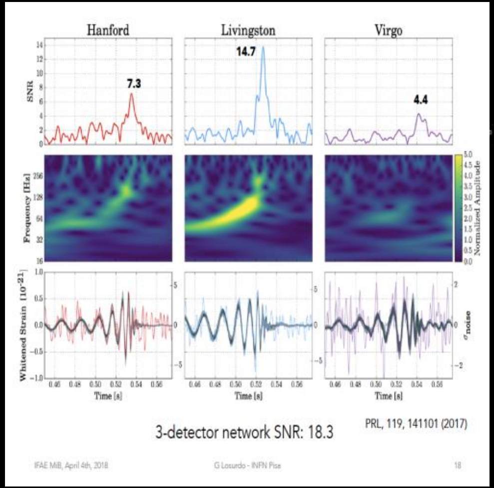
The first triple detection

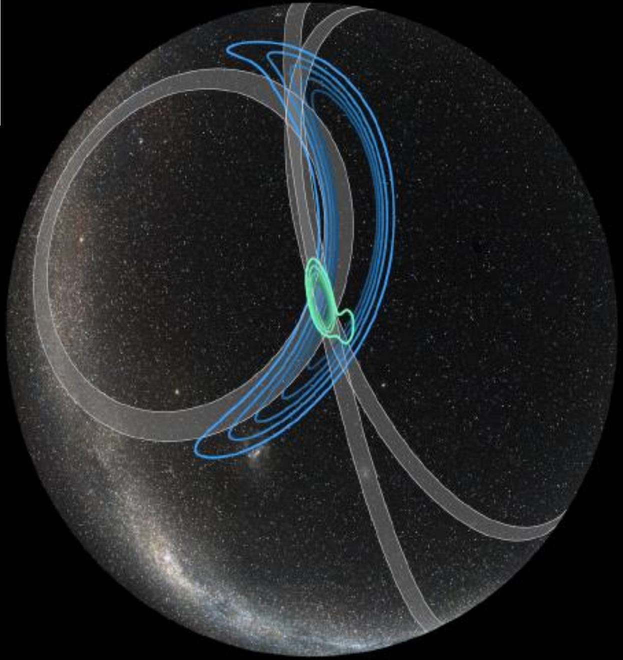


Credit: LIGO–Virgo



Virgo observed its first BBH coalescence ,GW170814



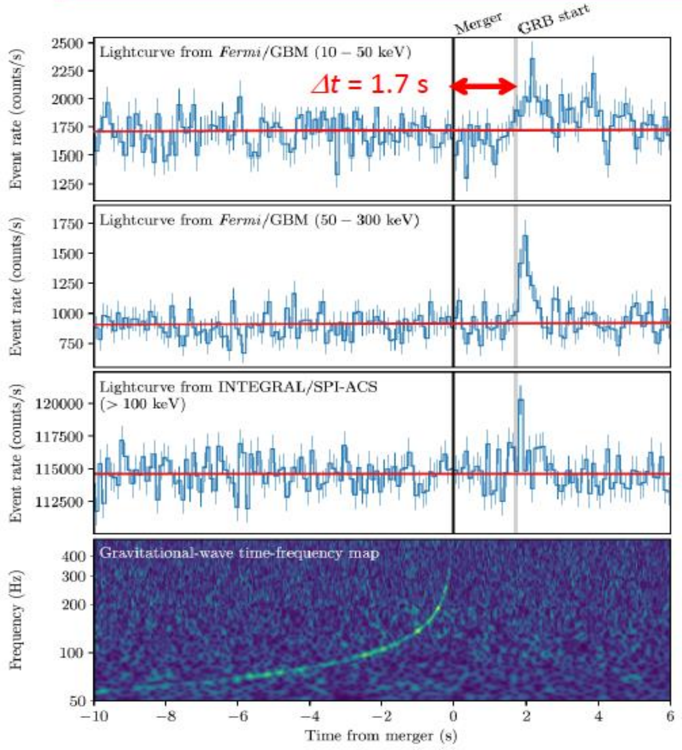


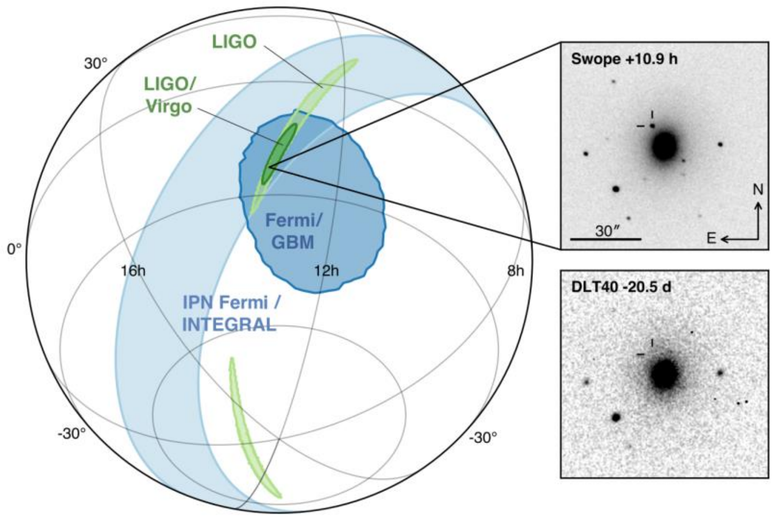
Credit: Leo Singer

LH 1160 square degrees

LHV 60 square degrees

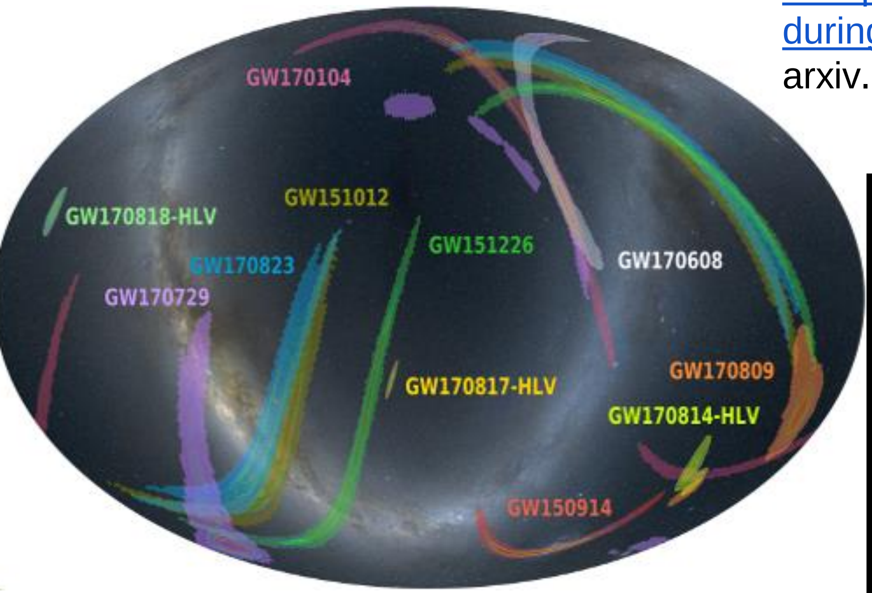
The MultiMessenger Astronomy



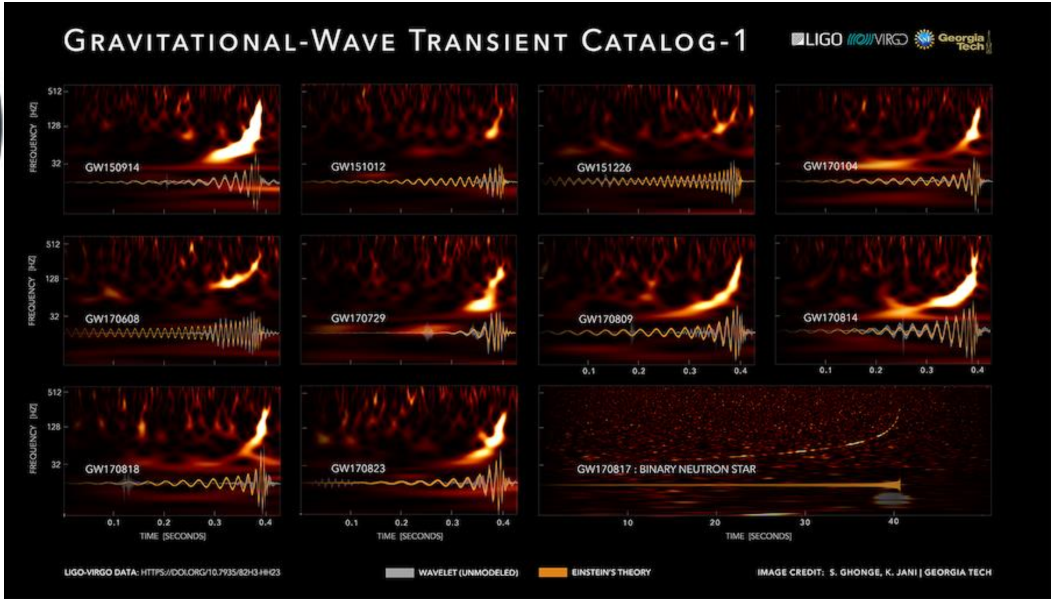


DOI:10.1103/PhysRevLett.119.161101.

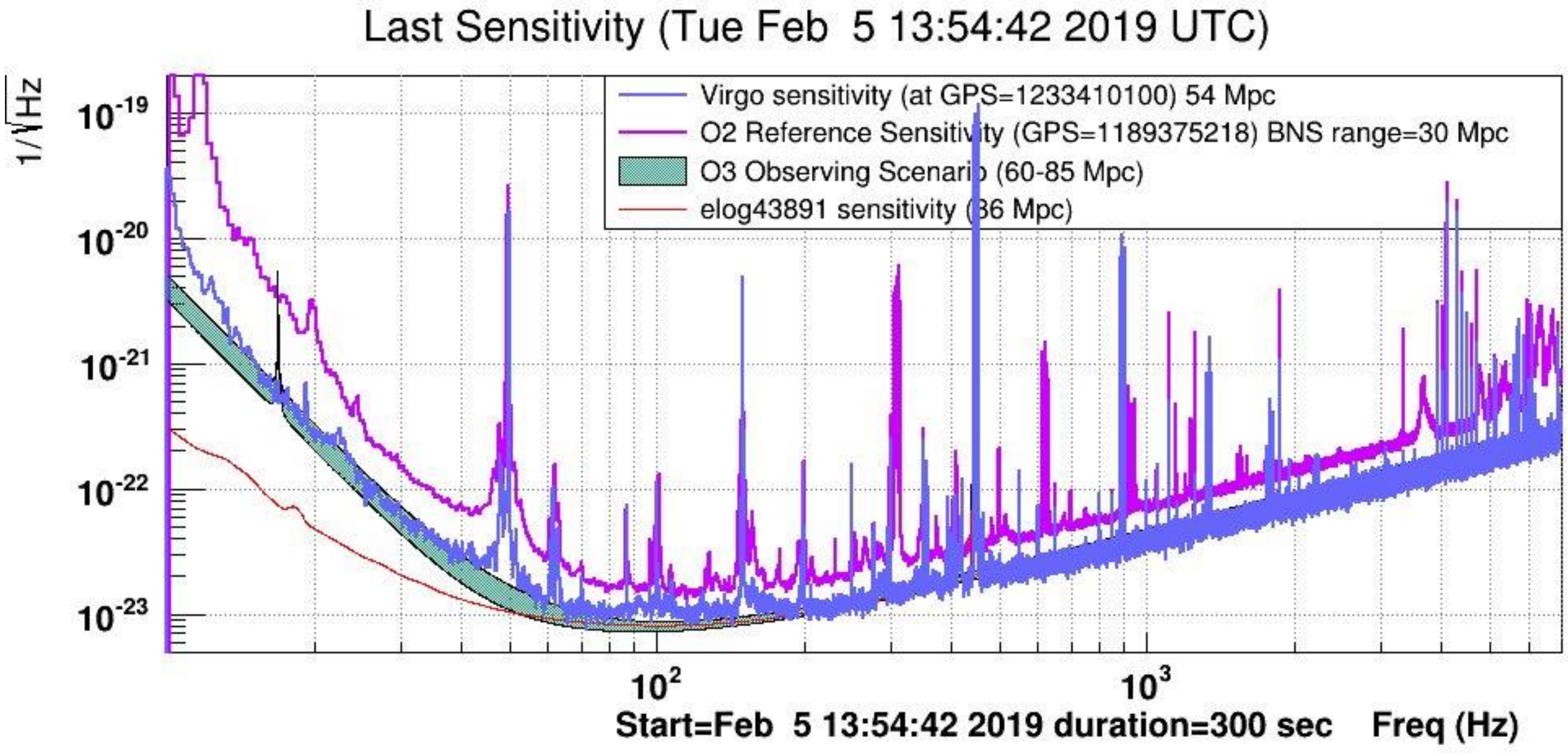
The first GW catalog

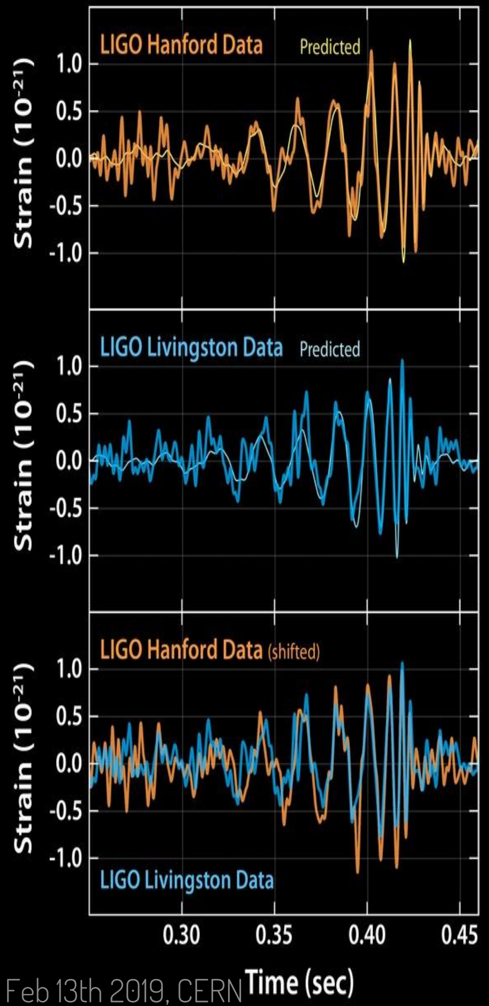


GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs
arxiv.org/abs/1811.12907

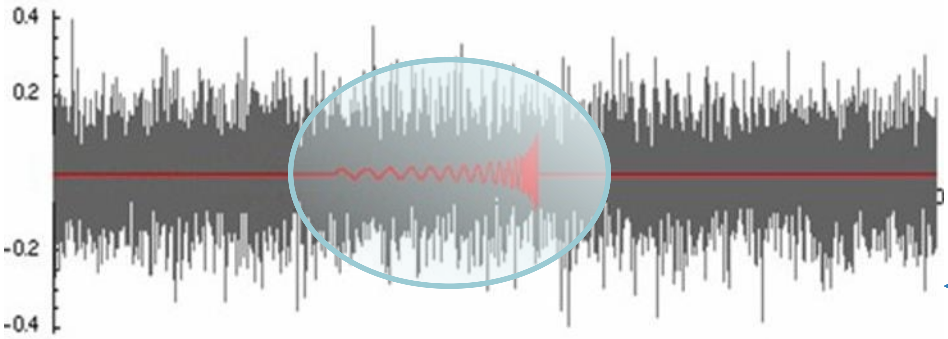


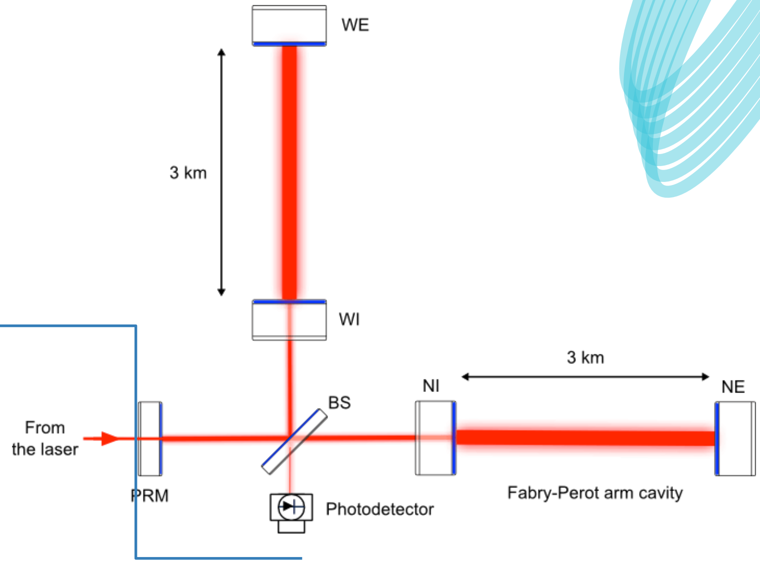
O3 is coming!


Last Sensitivity (Tue Feb 5 13:54:42 2019 UTC)
Virgo sensitivity (at GPS=1233410100) 54 Mpc
O2 Reference Sensitivity (GPS=1189375218) BNS range=30 Mpc
O3 Observing Scenario (60-85 Mpc)
elog43891 sensitivity (36 Mpc)
1/√Hz
10^-19
10^-20
10^-21
10^-22
10^-23
10^2
10^3
Start=Feb 5 13:54:42 2019 duration=300 sec
Freq (Hz)



Why Machine Learning in Gravitational Wave research





LIGO/Virgo data

are time series sequences… **noisy time series**
with low amplitude GW signal buried in

Elena Cuoco

Our "signals"

Astrophysical signals

Known GW signals

Compact coalescing binaries has known theoretical waveforms

Optimal filter: Matched filter

Too many templates to test

Unknown GW signals

Core collapse supernovae

No Optimal filter

Parameters estimation

Noise

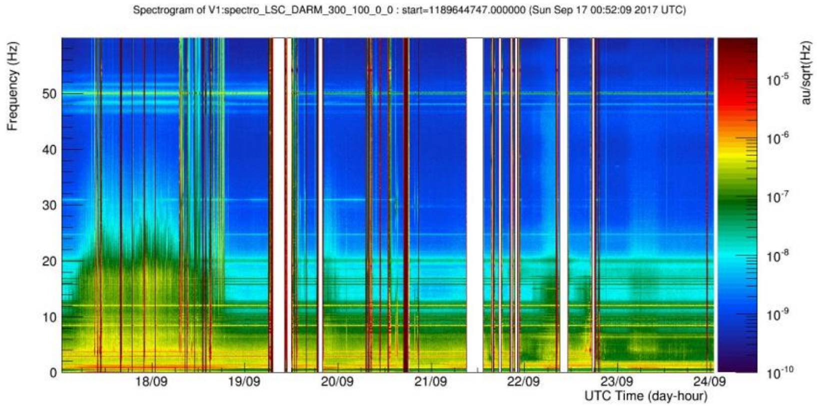
Moving lines

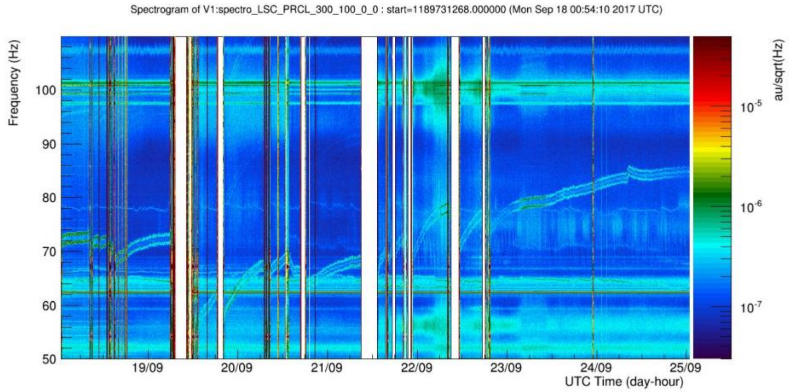
Broad band noise

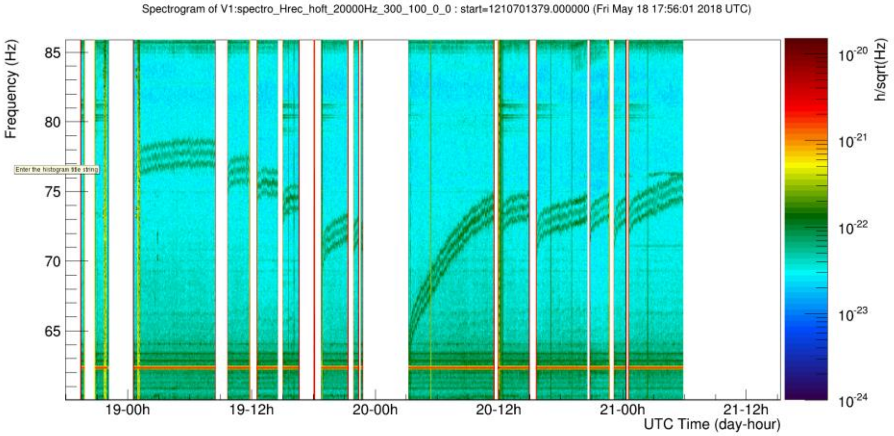
Glitch noise

"Pattern recognition" by visual inspection

Example of other noise signals

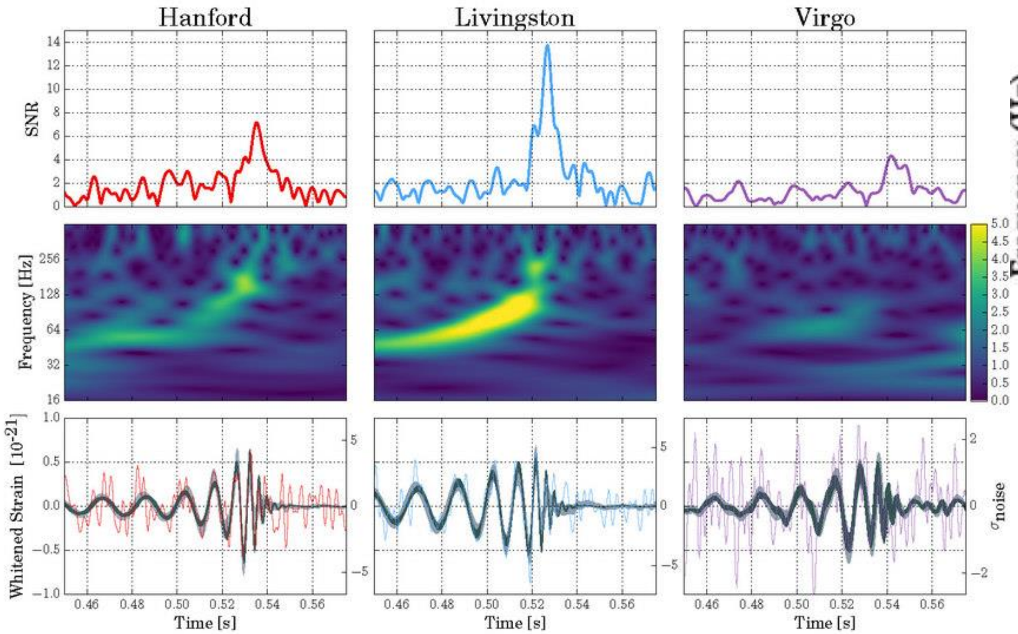




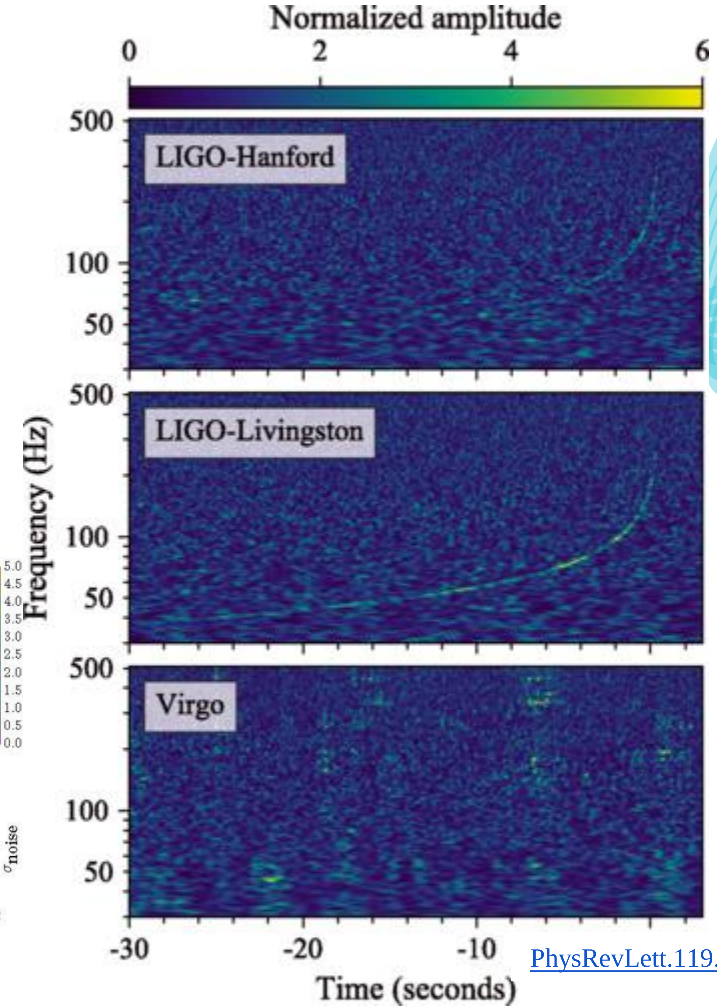


I. Fiori courtesy

Example of GW signals in Time-Frequency plots



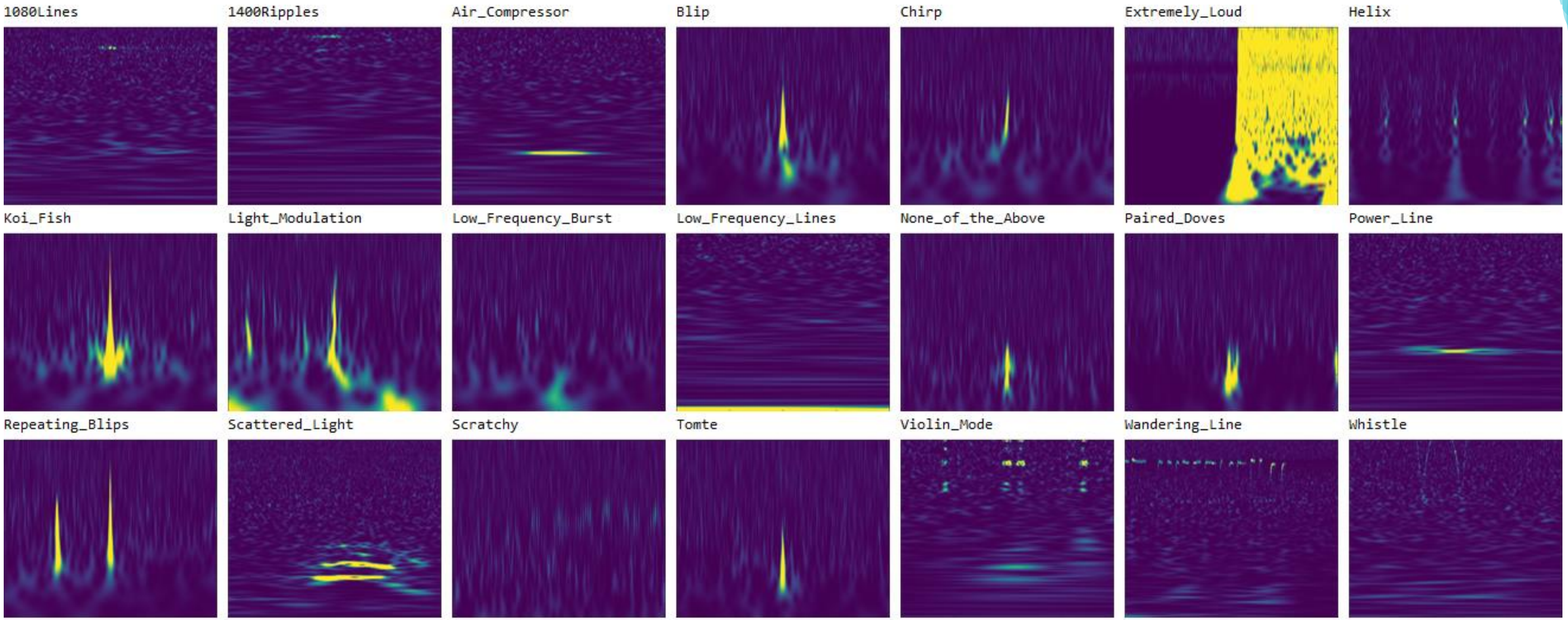
Phys. Rev. Lett., 119 (14), pp. 141101, 2017.



PhysRevLett.119.161101.

Example of Glitch signals

https://www.zooniverse.org/projects/zooniverse/gravity-spy



Gravity Spy, Zevin et al (2017)

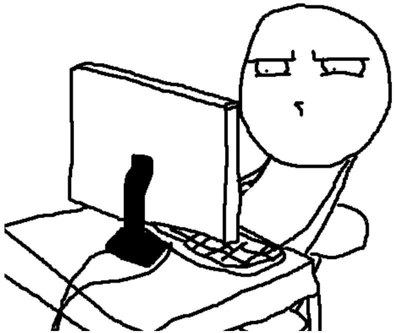
How Machine Learning can help

Data conditioning

- Identify Non linear noise coupling
- Use Deep Learning to remove noise
- Extract useful features to clean data

Signal Detection/Classification/PE

- A lot of fake signals due to noise
- Fast alert system
- Manage parameter estimation



Numbers about Virgo data

Data Stream Flux
- 50MB/s

Data on disk
- 1-3PB

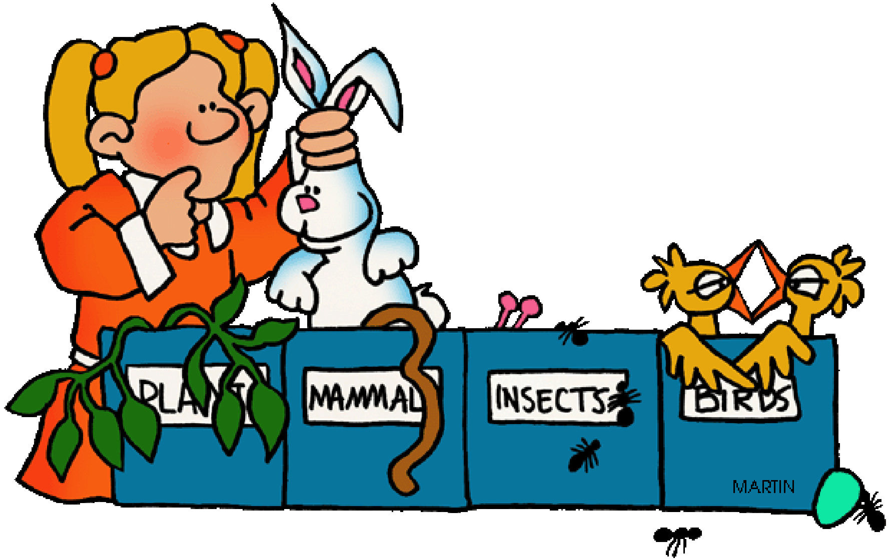
Number of events
- 1/week
- 1/day?

Number of glitches
- 1/sec
- 0.1/sec?

Should be analysed in less than 1min

Why Signal Classification?

- If we are able to classify the noise events, we can clean the data in a fast and clear way
- We can help commissioners
- We can identify glitch families

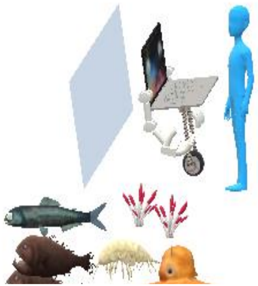


Machine learning models

Unsupervised

No label for the data

Semi-supervised

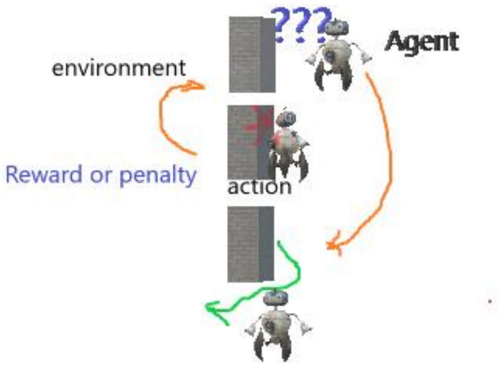


- Few labeled data
- A lot of not labeled data

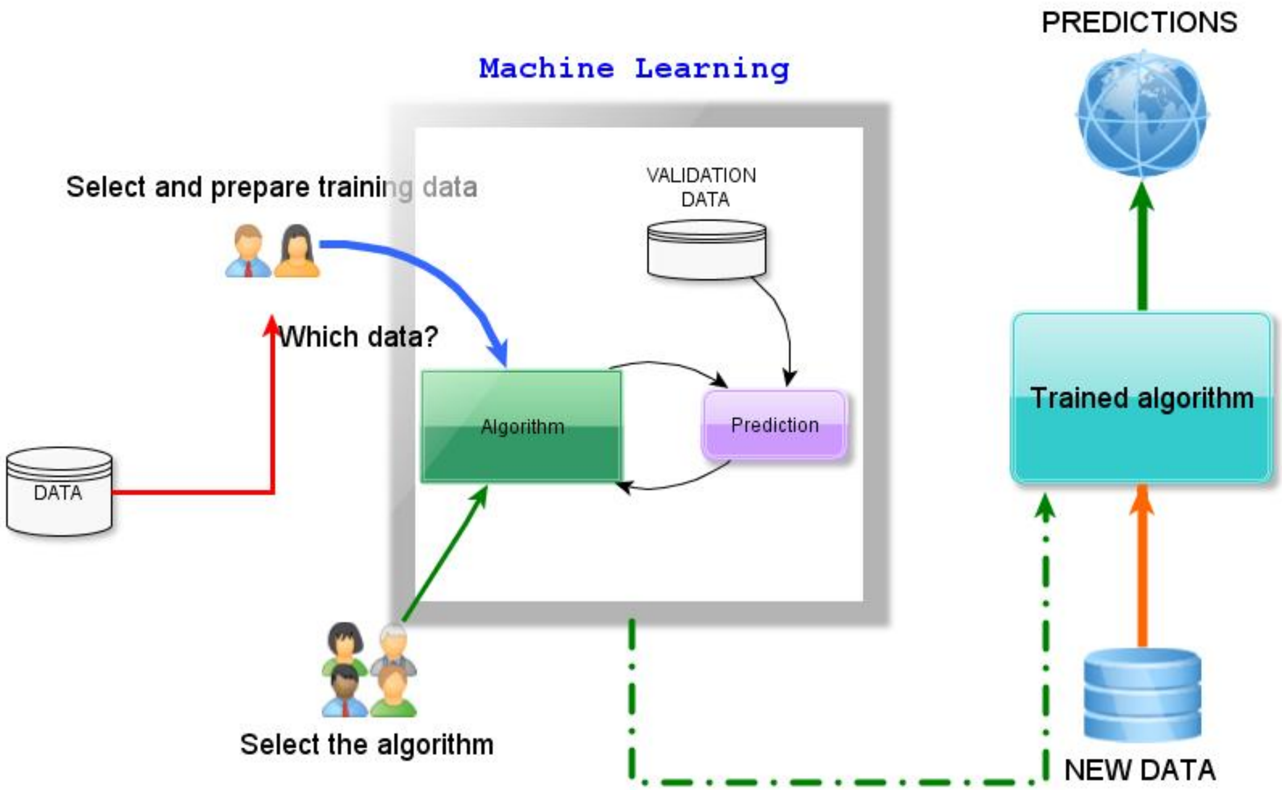
Supervised

Labeled training data

Reinforcement learning



Artificial Intelligence workflow



Elena Cuoco

What is going in the ML LIGO/Virgo group

136 LIGO/Virgo members

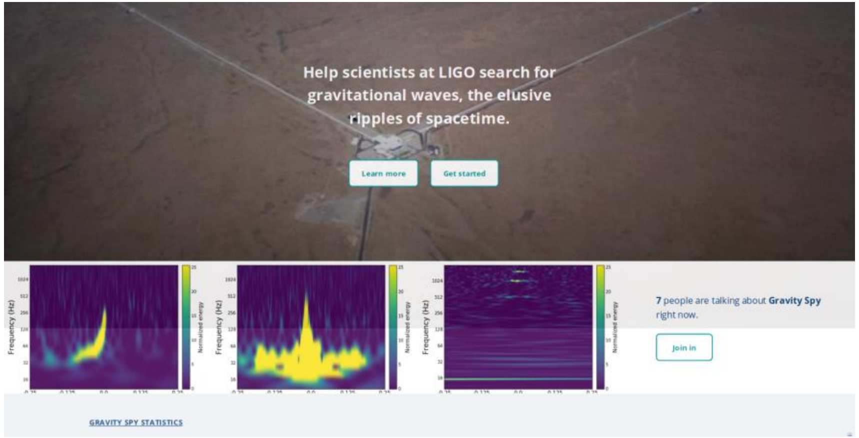
30 active projects

Example of interesting works

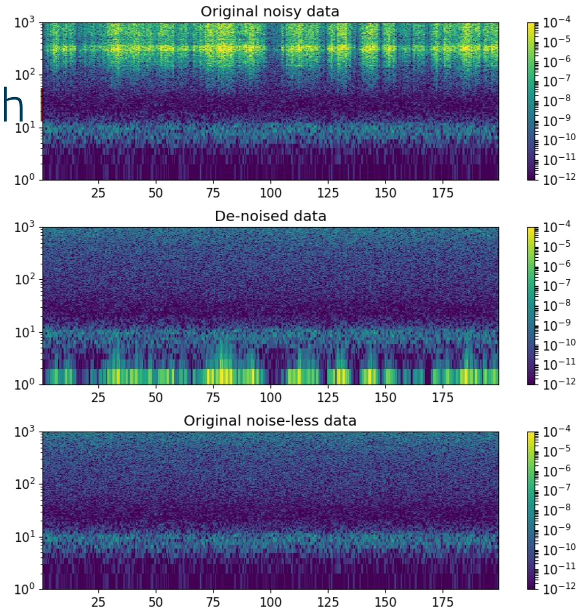
- Noise Removal

Non-linear and non-stationary noise subtraction with Learning

- Labelling glitches: Gravity Spy

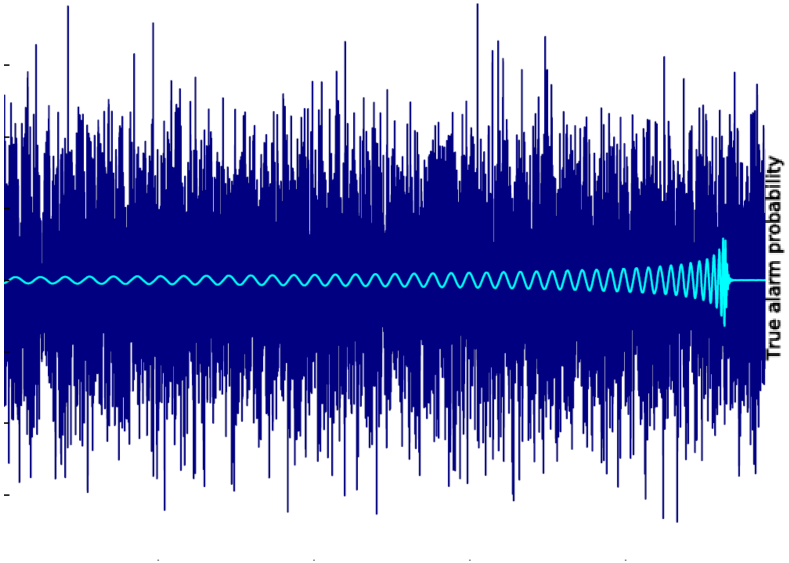


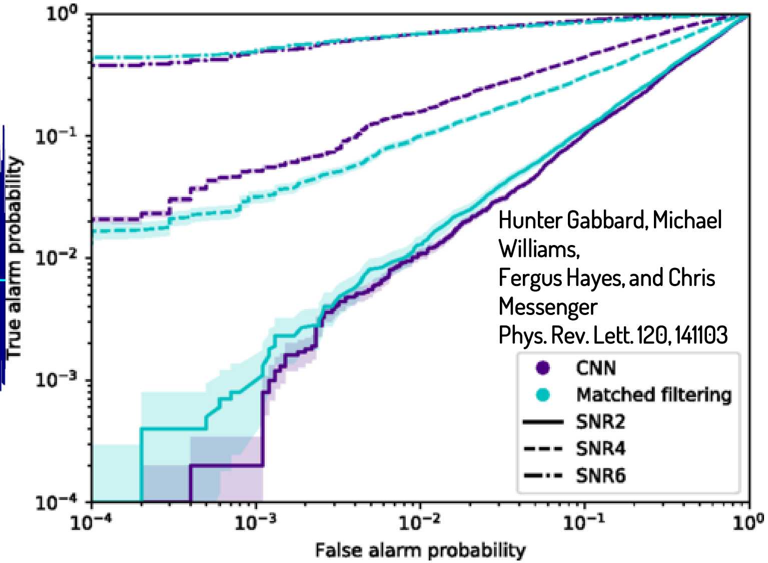
S. Coughlin courtesy



G. Vajente courtesy

Signal detection



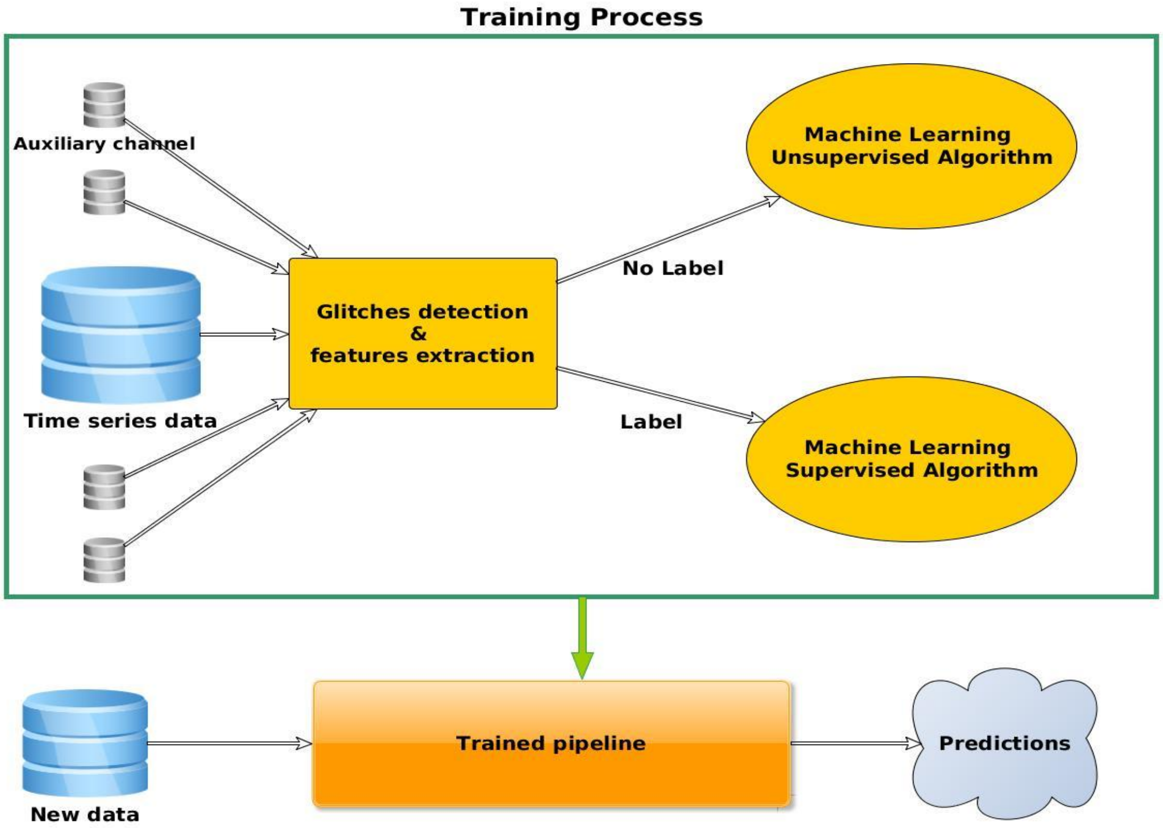


- Deep learning procedure requiring only the raw data time series as input with minimal signal pre-processing.
- Performance similar to Optimal Wiener Filter

Glitches classification efforts in LIGO/Virgo Community

- Gravity Spy (M. Zevin,S. Coughlin,J. R. Smith, A. Lundgren, D. Macleod, V. Kalogera)
- Wavefier(E. Cuoco et al.)
- WDFX (E. Cuoco, M. Razzano, A. Utina)
- Karoo GP (K. Staats, M. Cavaglià)
- Wavelet-DBNN (N. Mukund S. Abraham S. Mitra et al)
- ImageGlitch CNN (M. Razzano, E. Cuoco)
- Low latency transient detection and classification (I. Pinto, V. Pierro, L. Troiano, E. Mejuto-Villa, V. Matta, P. Addesso)
- Deep Transfer Learning (Daniel George, Hongyu Shen, E.A. Huerta)
- Gstlal-iDQ (P. Godwin, R. Essick, D. Meacher, S. Chamberlain, C. Hanna, E. Katsavounidis, L. Wade, M. Wade, D. Moffa, K. Rose)
- New ranking statistic for gstlal (K. Kim, T.G.F. Li, R.K.-L. Lo, S. Sachdev, R.S.H. Yuen)
- RGB image SN CNN (P. Astone, S. Frasca, C. Palomba, F. Ricci, M. Drago, I. Di Palma, F. Muciaccia, Pablo Cerda-Duran)

Glitch classification strategy for GW detectors



Two different approaches

- Images

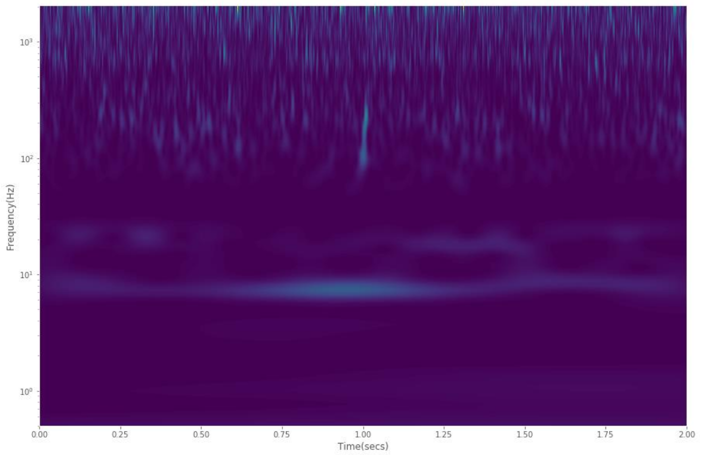
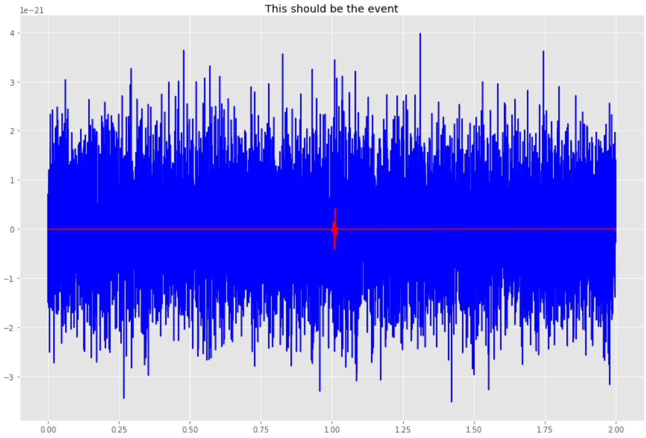


Image-based deep learning for classification of noise transients in gravitational wave detectors, Massimiliano Razzano**, Elena Cuoco**, Class.Quant.Grav. 35 (2018) no.9, 095016

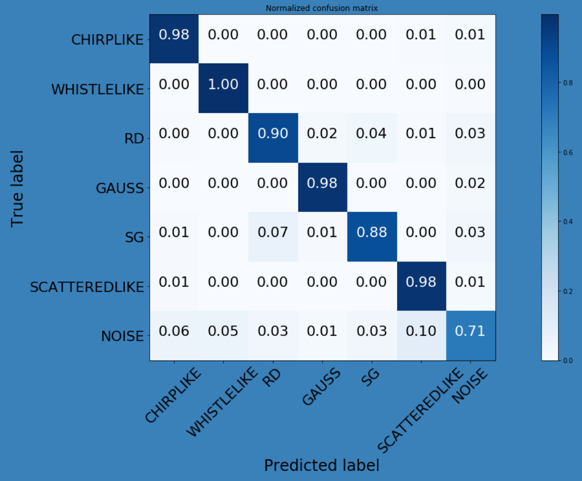
- Time series



Wavelet-based Classification of Transient Signals for Gravitational Wave Detectors**, Elena Cuoco**, Massimiliano Razzano and Andrei Utina, #1570436751 accepted reviewed paper at EUSIPCO2018

Glitches classifcation

- Application on Simulated data
- Application on Real Data
- Time-series (Wavelet) based classification
- Image based classification with Deep Learning



Test on simulated data sets

To test the pipeline, we prepared ad-hoc simulations

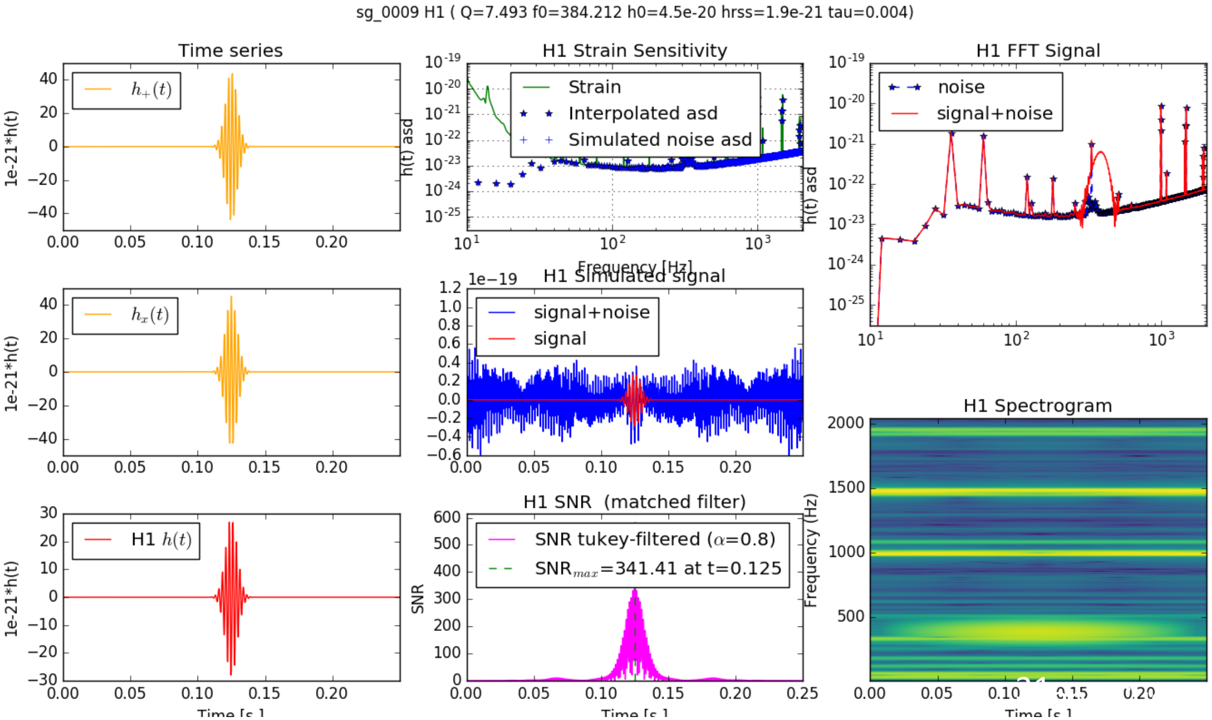
Simulate colored noise using public H1 sensitivity curve

Add 6 different classes of glitch shapes

Data simulation

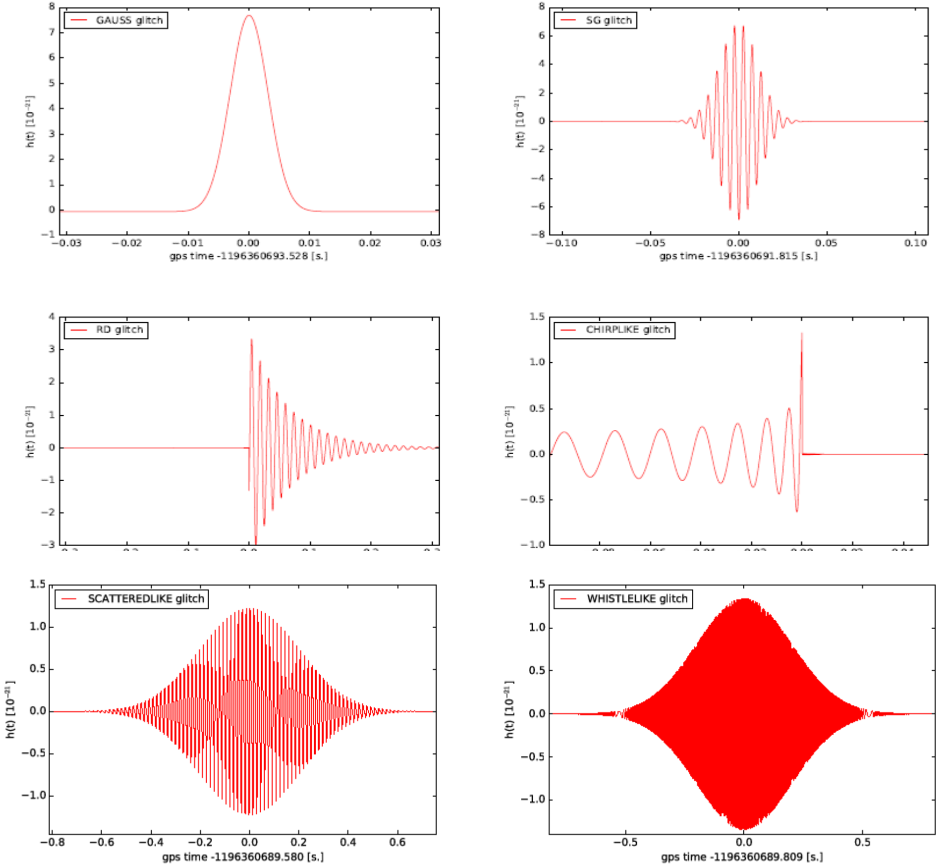
- Ad hoc simulations for tests (e.g. Powell+2015)
- Simulate colored noise using public sensitivity curve
- 6 classes of glitch shapes (+ NOISE one to check detection)

Razzano's courtesy



Example of H1 simulation

Simulated signal families



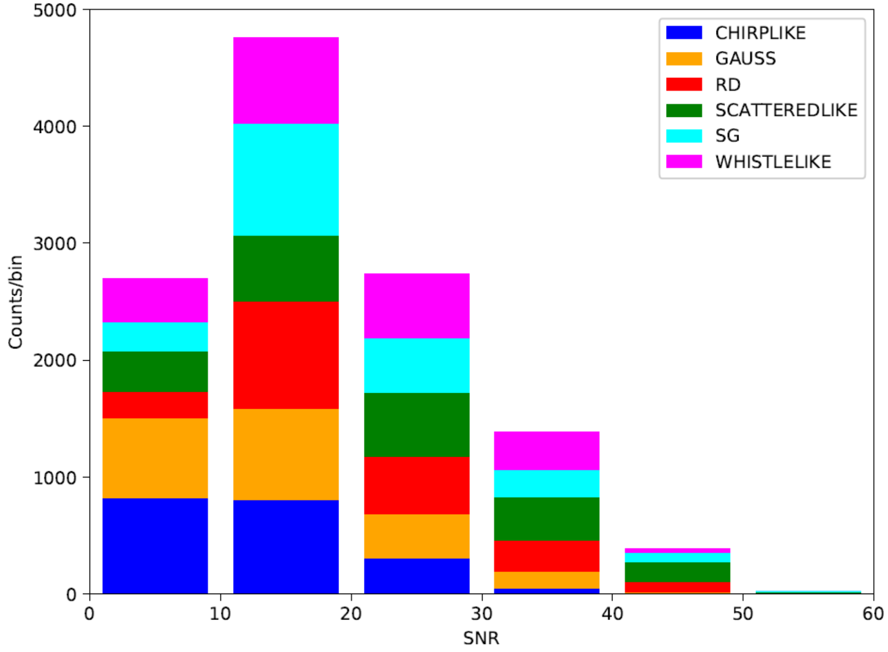
Waveform
Gaussian
Sine-Gaussian
Ring-Down
Chirp-like
Scattered-like
Whistle-like
NOISE (random)

To show the glitch time-series here we don't show the noise contribution

Razzano M., Cuoco E. CQG-104381.R3

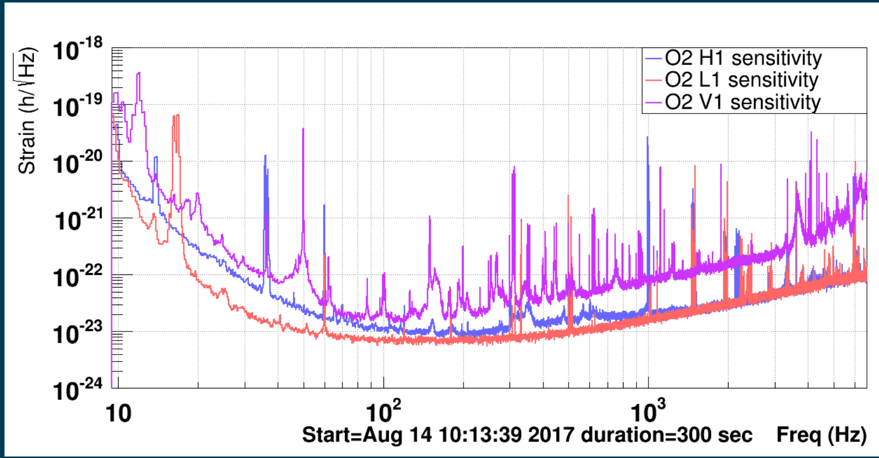
Signal distribution

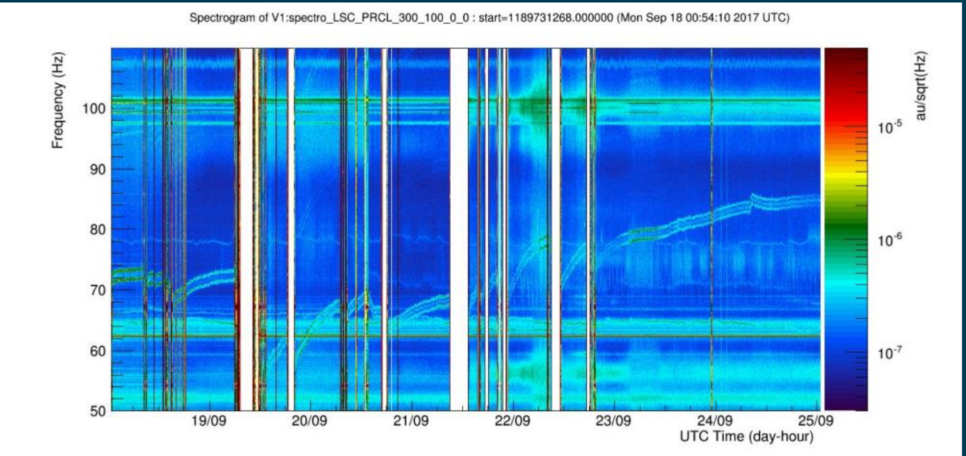
- Simulated time series with 8kHz sampling rate
- Glitches distributed with Poisson statistics m=0.5 Hz
- 2000 glitches per each family
- Glitch parameters are varied randomly to achieve various shapes and Signal-To-Noise ratio



Data preprocessing

- Many spectral features
- Non stationary and non linear noise





Whitening in time domain

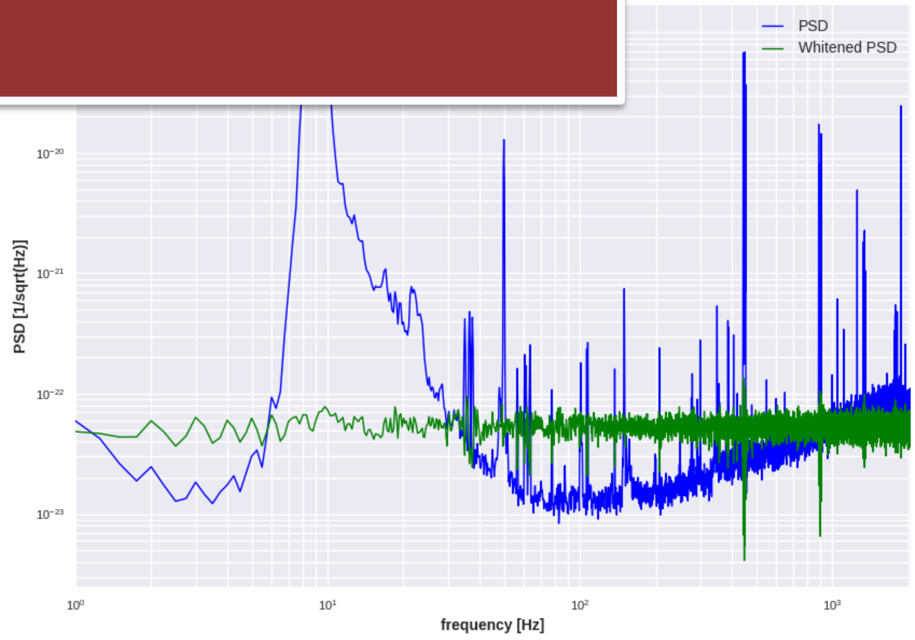
We need parametric modeling

It can be use

line application

It can be implemented for non stationary noise

It can catch the autocorrelation function to larger lags



AR parametric modeling

An AutoRegressive process is governed by this relation

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + w[n],$$

and its PSD for a process of order P is given by

$$P_{AR}(f) = \frac{\sigma^2}{|1 + \sum_{k=1}^{P} a_k \exp(-i2\pi kf)|^2}$$

Kay S 1988 Modern spectral estimation: Theory and Application Prentice Hall Englewood Cliffs

Advantages of AR modeling

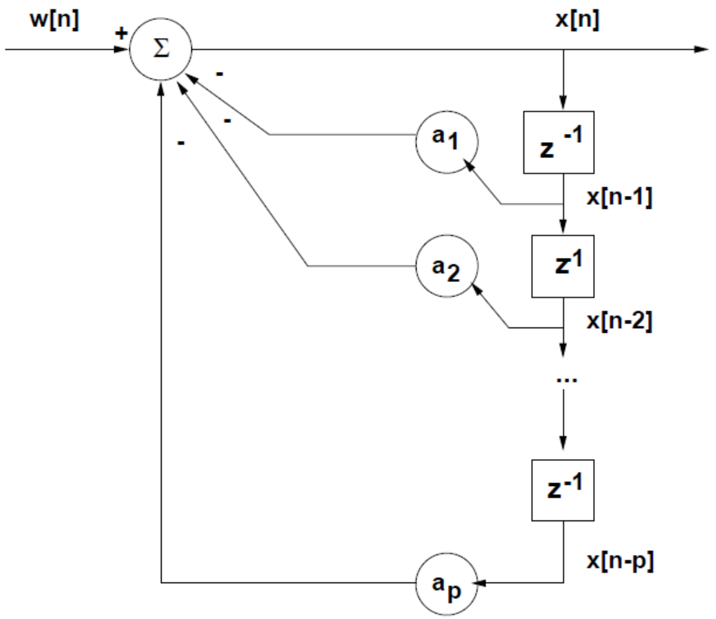
- Stable and causal filter: same solution of **linear predictor filter**

$$\hat{x}[n] = \sum_{k=1}^{P} w_k x[n-k] .$$

$$e[n] = x[n] - \hat{x}[n]$$

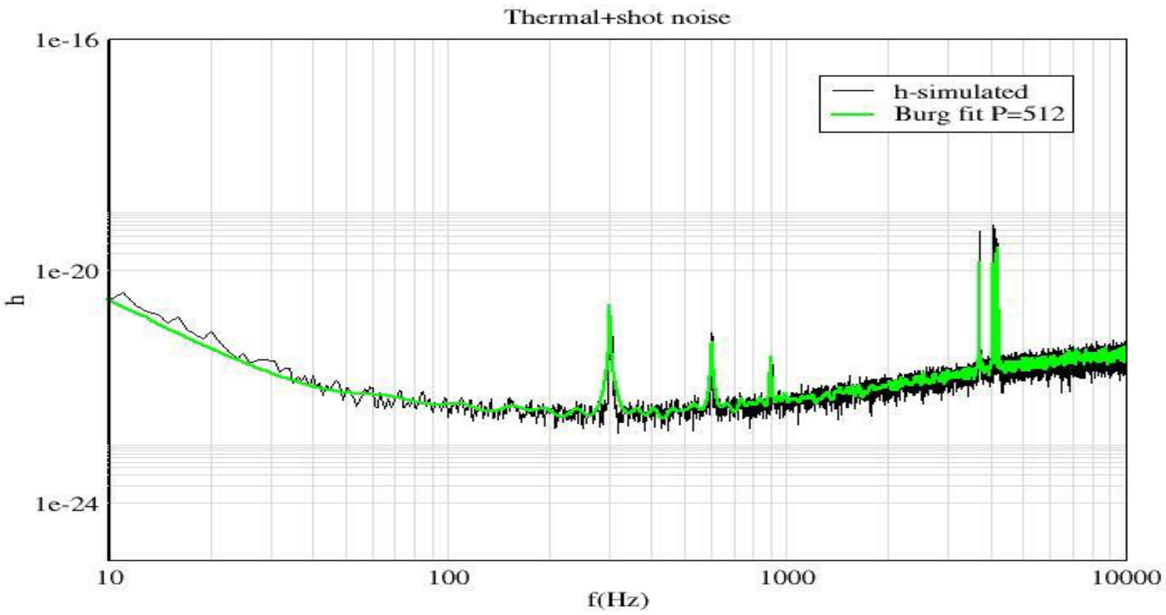
$$\varepsilon_{min} = r_{xx}[0] - \sum_{k=1}^{P} w_k r_{xx}[-k] ,$$

$$\begin{aligned} w_k &= -a_k \\ \varepsilon_{min} &= \sigma^2 \end{aligned}$$



Wiener-Hopf equations

PSD AR(P) Fit



Cuoco et al. Class.Quant.Grav. 18 (2001) 1727-1752 and
Cuoco et al.Phys.Rev.D64:122002,2001

Lattice Filter

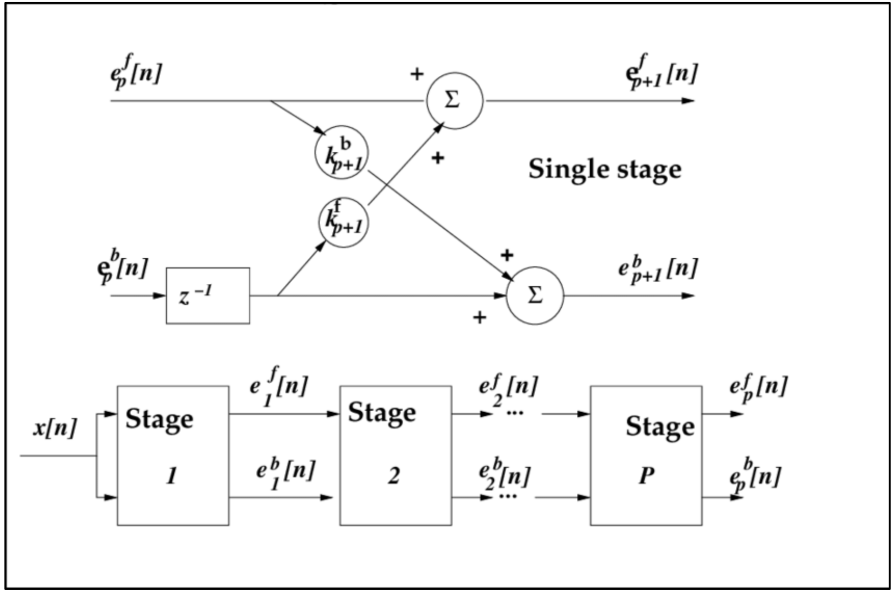
The Least Squares based methods build their cost function using all the information contained in the error function at each step, writing it as the sum of the error at each step up to the iteration n

Forgetting factor

$$\epsilon[n] = \sum_{1}^{n} \lambda^{n-1} e^{2}(i|n)$$

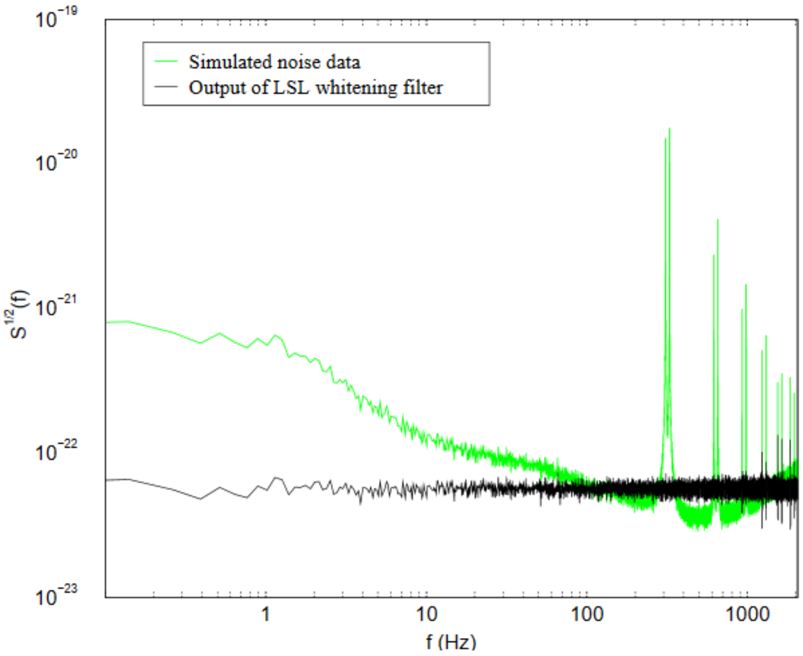
$$e(i|n) = d[i] - \sum_{k=1}^{N} x_{i-k} w_k[n],$$

Desired signal



Adaptive whitening using Lattice Filter

- If $\lambda = 1$ we are in the stationary data
- If $0 < \lambda < 1$ we can follow non stationary noise
- The Least Square Lattice filter is a modular filter with a computational cost proportional to the order P



Whitening in time domain

Static whitening

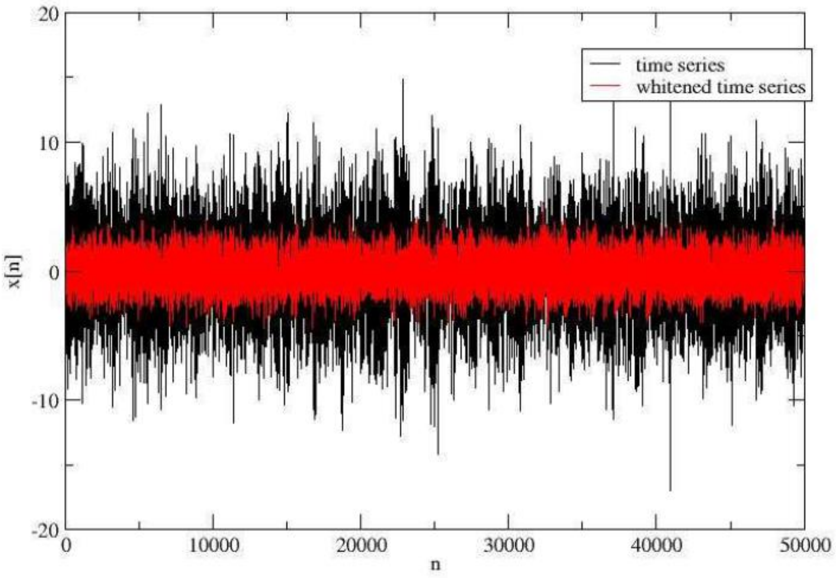
- We estimate the AR and reflection coefficients in a first part of the data
- We assume the data are stationary
- We setup a Lattice structure to run on line the whitening filter in time domain.

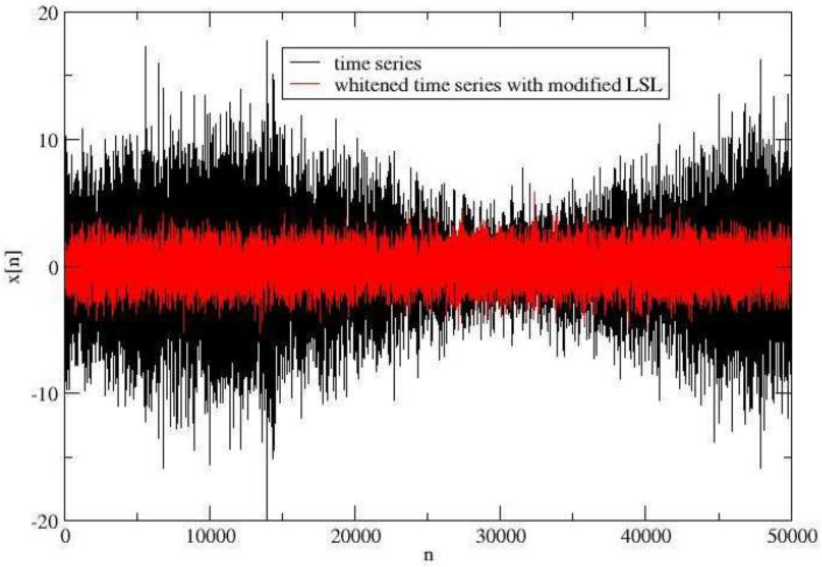
Adaptive whitening

- We make only a guess of the rmse
- We start estimating the reflection coefficients while acquiring data
- We use the forgetting factor to follow and remove the slow non stationary noise

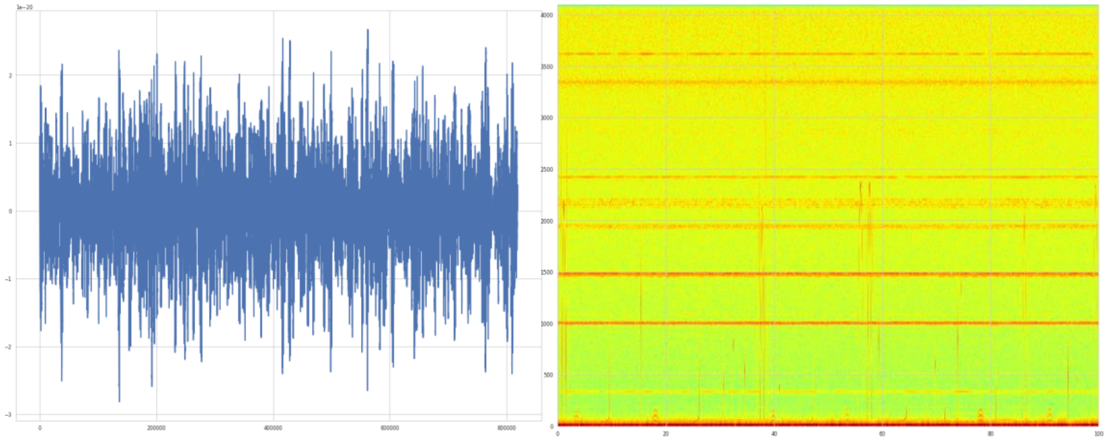
Whitened data in time domain

Example on simulated data

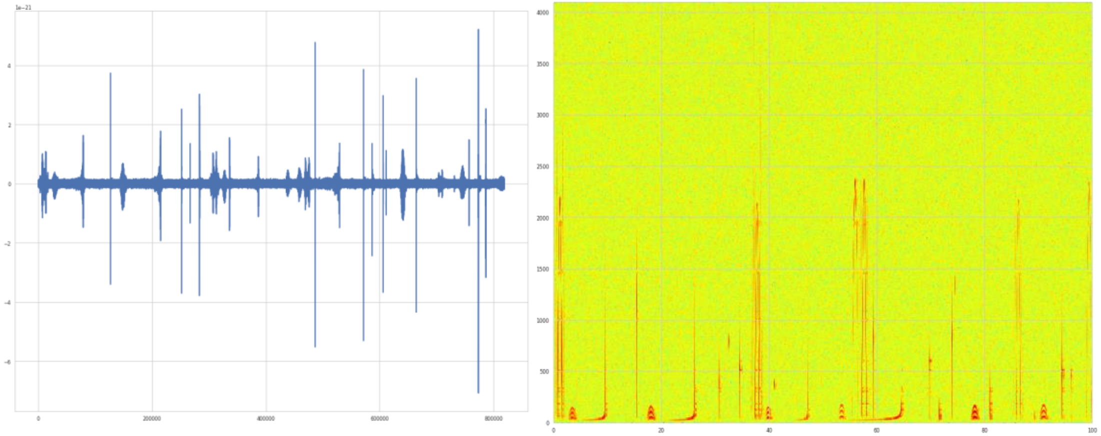




Signals in whitened data



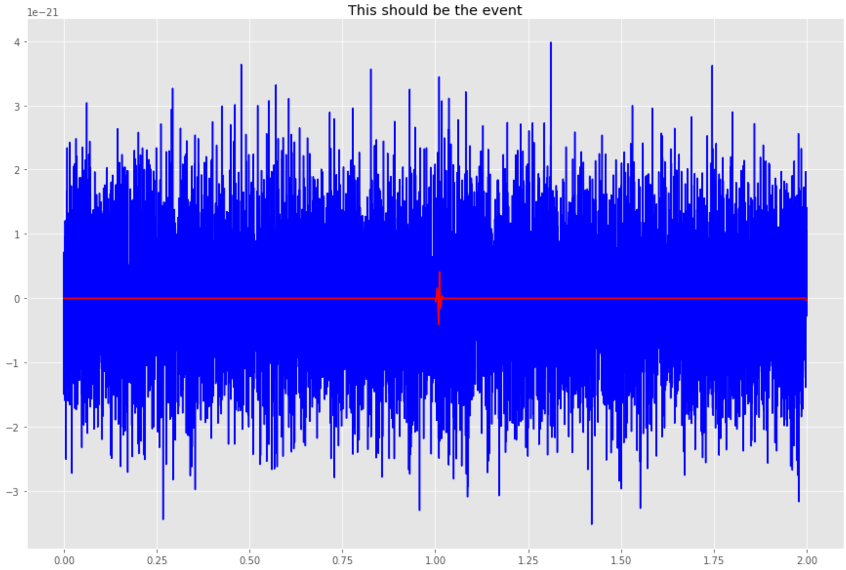
Not Whitened



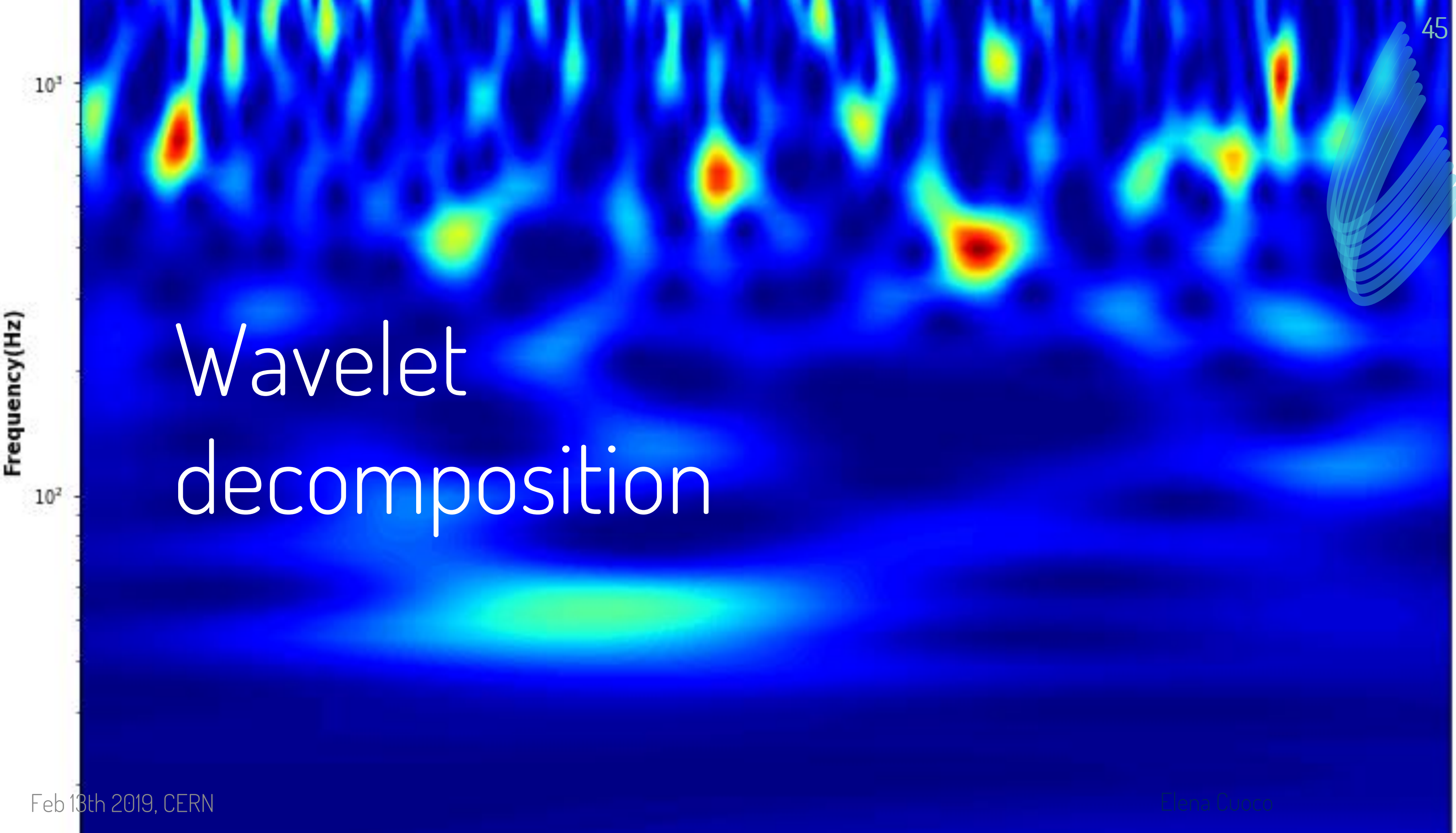
Whitened

Wavelet based classification

- Time series

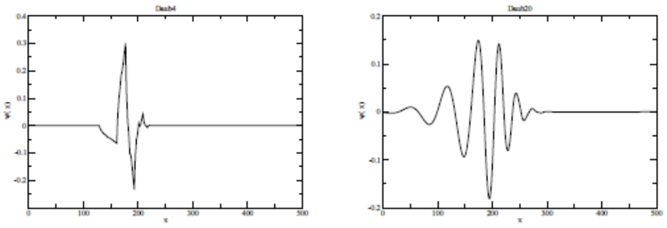


Wavelet decomposition



Wavelet decomposition of time series

The wavelet transform replaces the Fourier transform sinusoidal waves by a family generated by translations and dilations of a window called a wavelet.



$$Wf(a,b) = < f, \psi_{a,b} > = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{b}} \psi^*\left(\frac{t-a}{b}\right) \, dt$$

Wavelet denoising

$$x_i = h_i + n_i \quad i = 0, 1, \ldots N-1$$

$$W(x) = W(h) + W(n)$$

Wavelet transform

$$t = \sqrt{2 \log N}\hat{\sigma}$$

Local noise

$$\hat{h} = W^{-1}(T(Wx))$$

Threshold function

Dohone and Johnston proposed two different thresholding strategy: the soft thresholding and the hard thresholding. Given a threshold t and w the wavelet coefficient, the hard threshold for the signal is w if $|w| > t$, and is 0 if $|w| < t$. The soft threshold for the signal is $sign(w)(|w| - t)$ if $|w| > t$ and is 0 if $|w| < t$.

Wavelet Detection filter as Event Trigger Generator

$$E_s = \sqrt{\sum_{k,j} w_{k,j}^2}$$

- Select highest values

$\propto$ Energy of the signal

$$SNR = \frac{E_s}{\sigma}$$

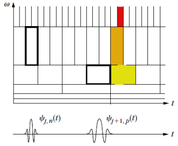
- Reconstruct a proto-SNR

$\propto$ SNR of the signal

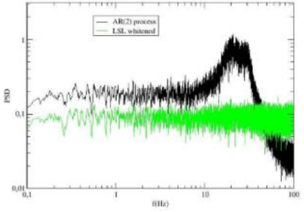
Wavelet Detection Filter (WDF) workflow

$$x_i = h_i + n_i, \quad i = 0, 1, \ldots N-1,$$

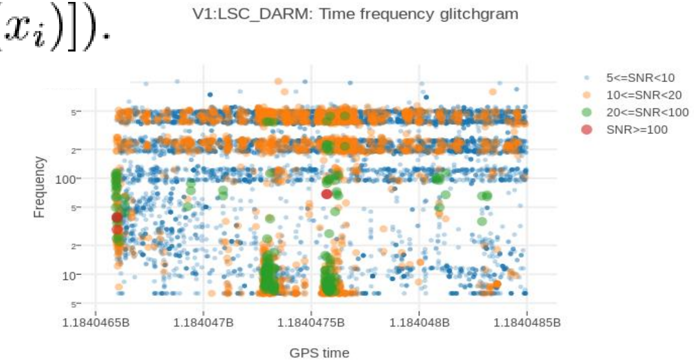
$$Wf(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{b}} \psi^* \left(\frac{t-a}{b} \right) \; dt,$$



Data → Whitening in time domain → Wavelet transform → De-noising → Parameter estimation → Trigger list

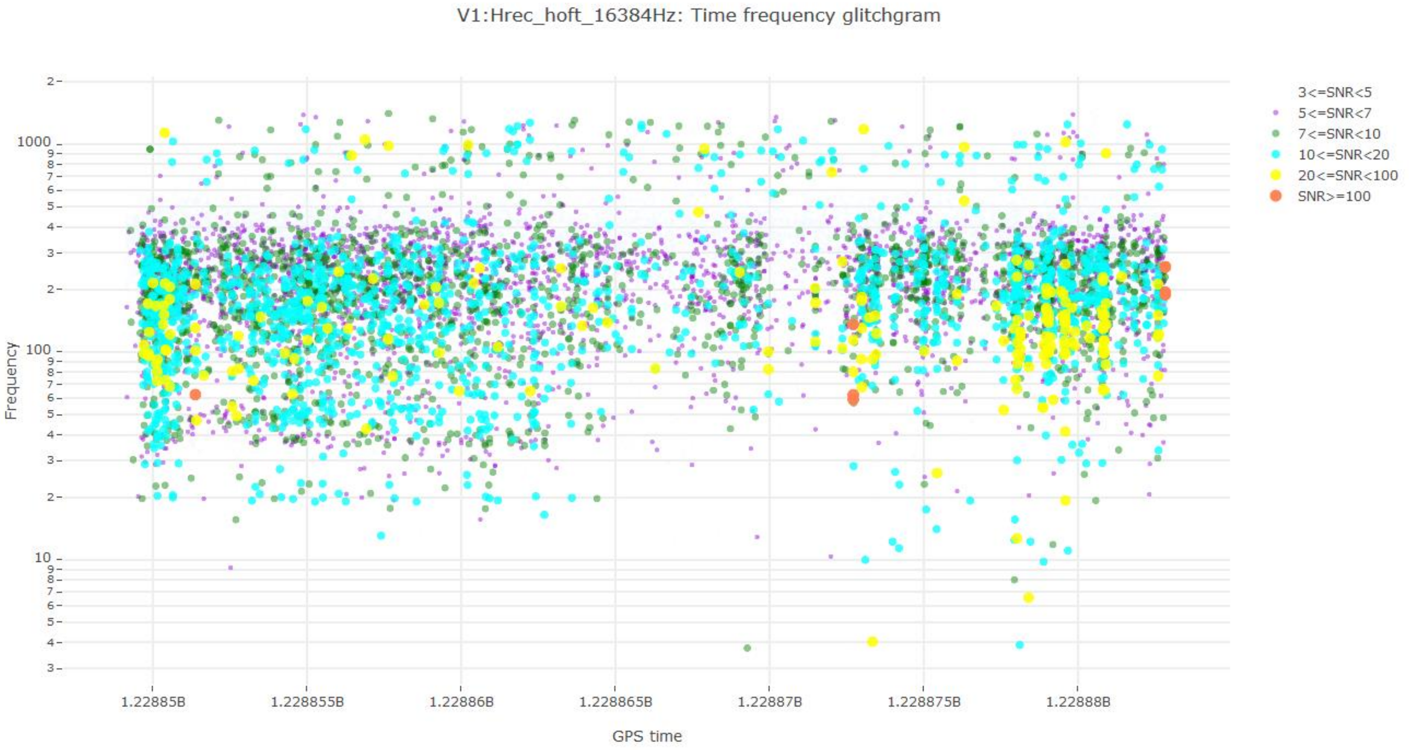


$$\hat{h}_i = W^{-1}(t[W(x_i)]).$$



Glitchgram

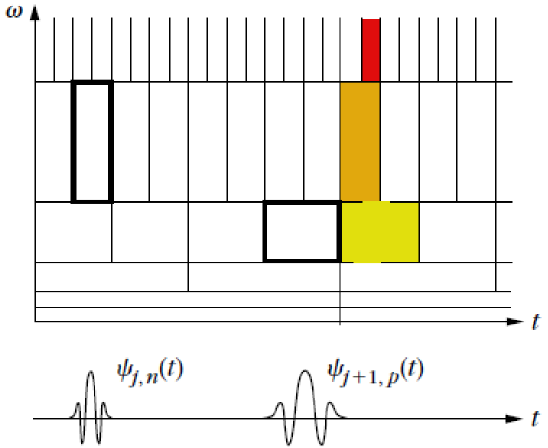
Time-Frequency distribution by SNR slice



Elena Cuoco

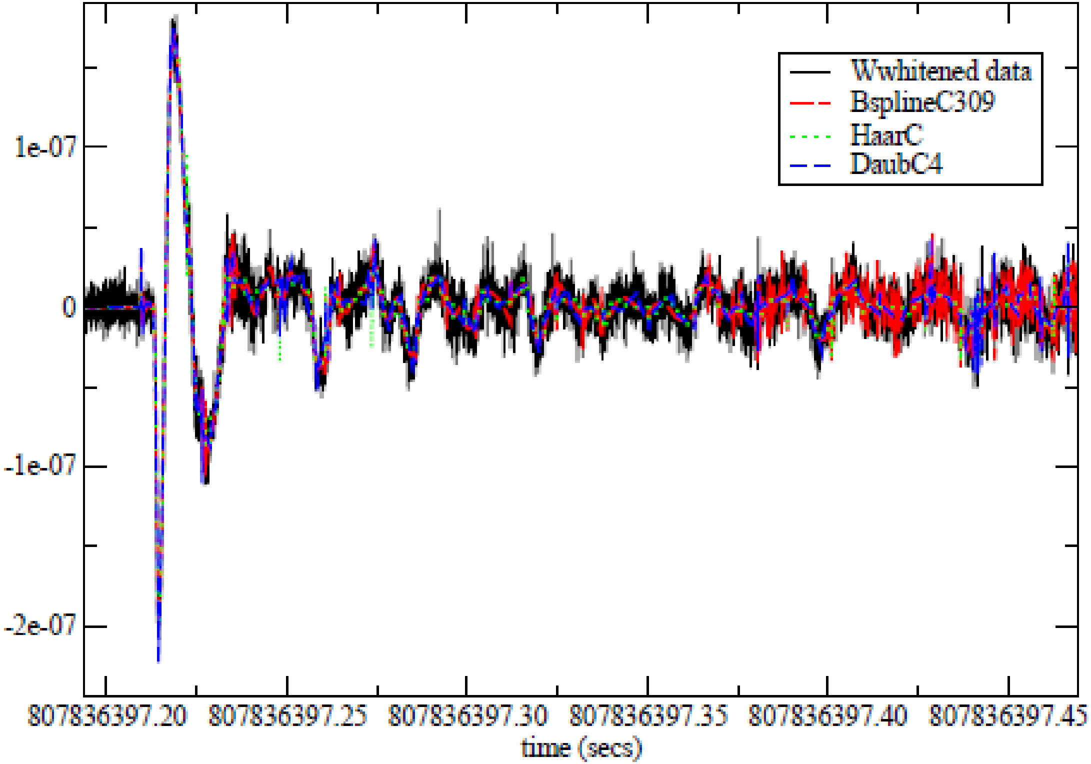
WDF waveform extraction

- Wavelet transform in the selected window size
- Retain only coefficients above a fixed threshod (Donoho-Johnston denoise method)
- Create a metrics for the energy using the selected coefficients and give back the trigger with all the wavelet coefficients.
- In the wavelet plane, select the highest values to build the event
- Inverse wavelet transform
- Estimate mean and max frequency and snr max of the cleaned event

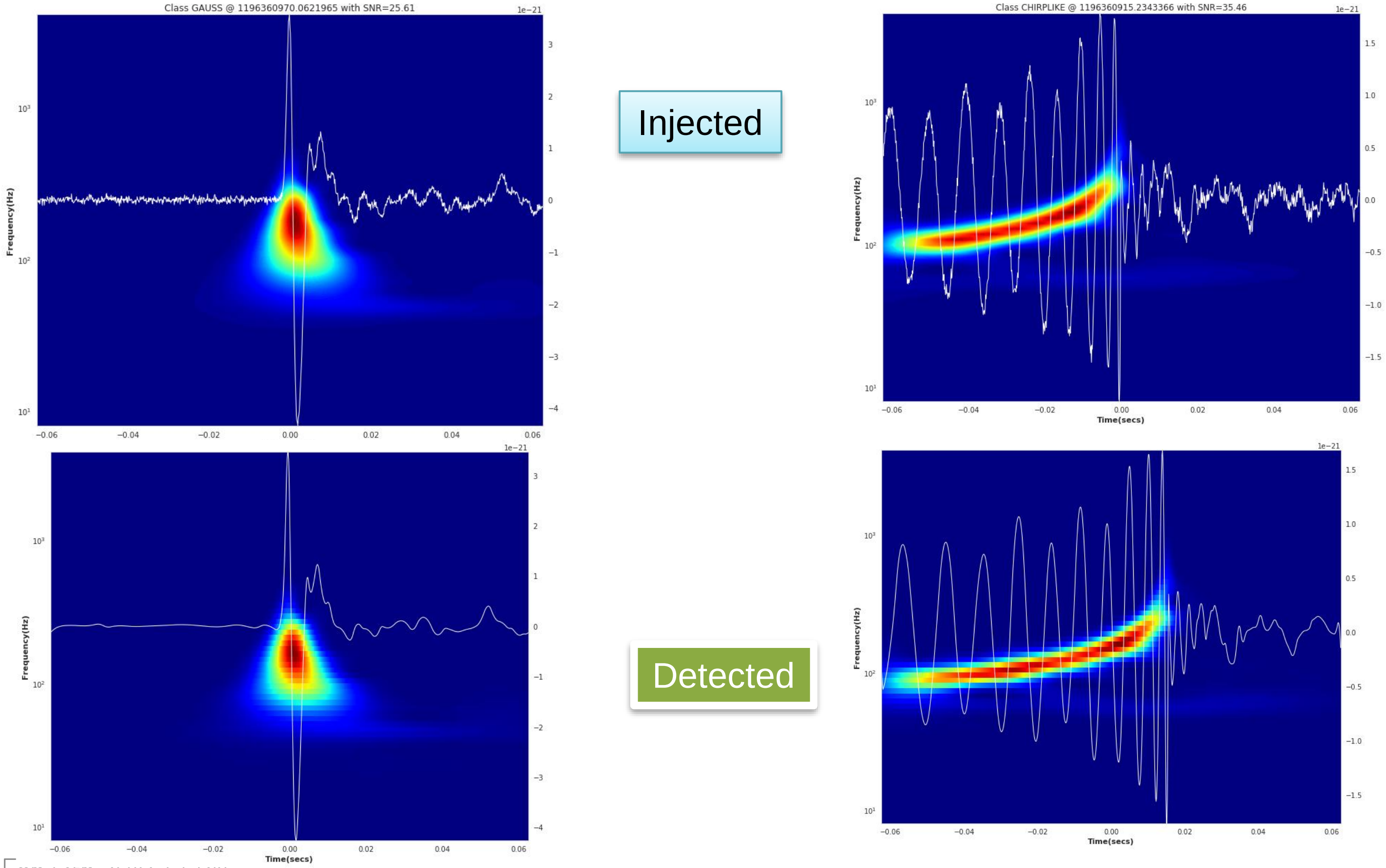


Gps, duration, snr, snr@max, freq_mean, freq@max, wavelet type triggered + corresponding wavelets coefficients.

Waveform reconstruction

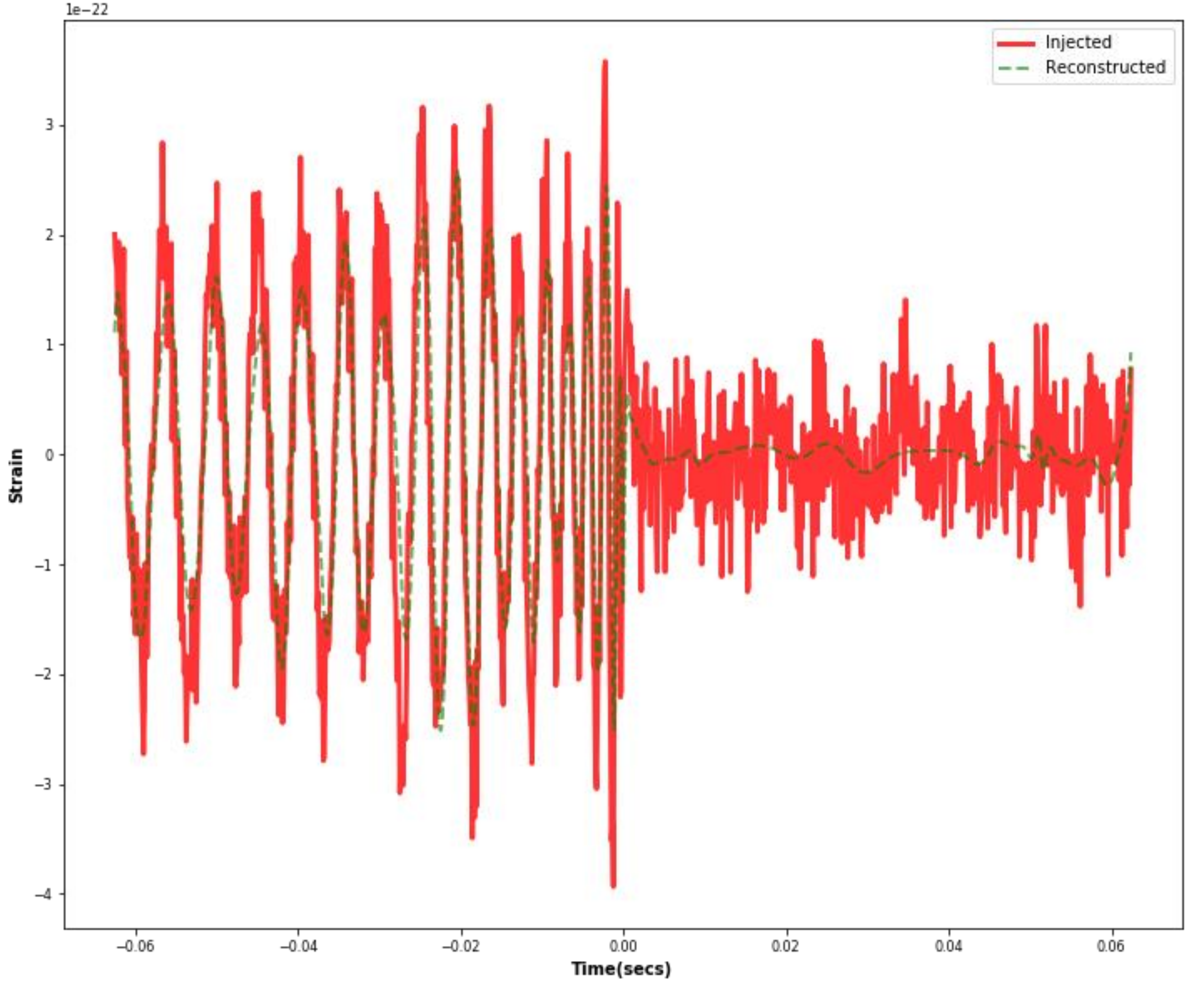


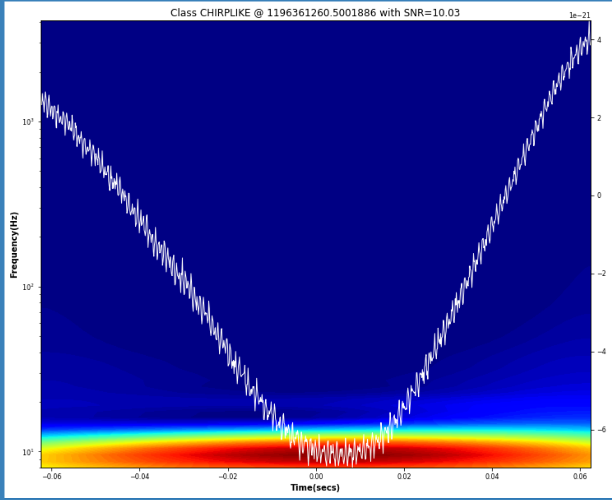
Waveform reconstruction: example



Elena Cuoco

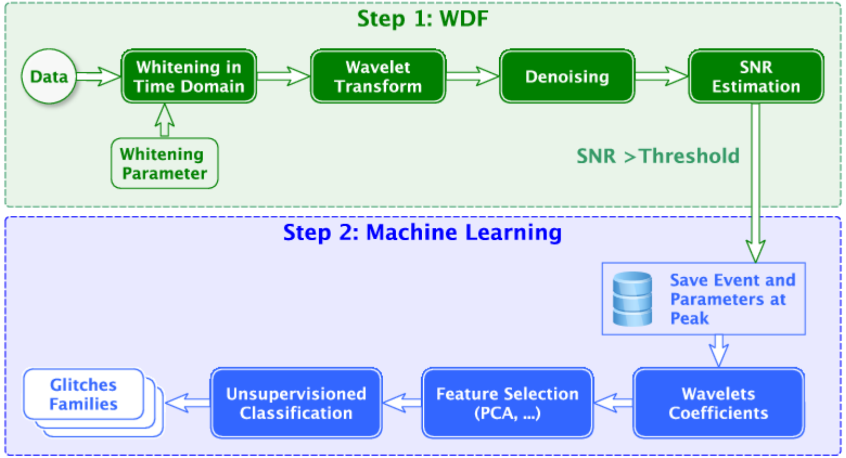
Injection and Reconstruction in perfect match



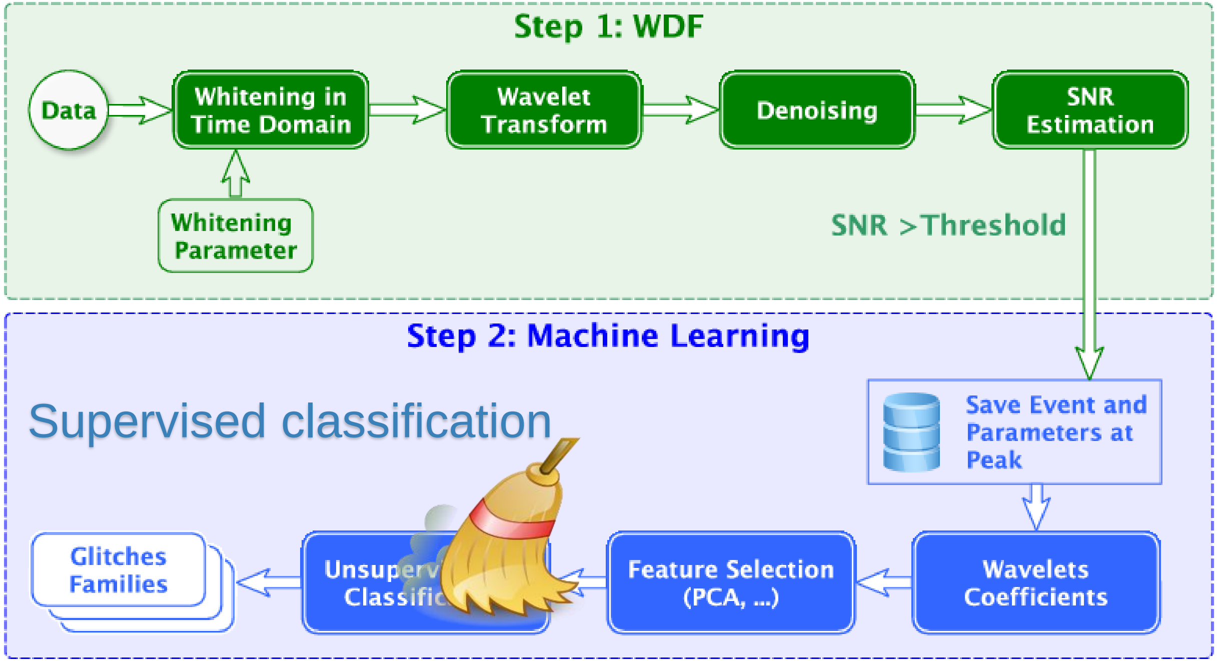


Glitch classification

- Unsupervised on Simulated data:
  - Classification methods for noise transients in advanced gravitational-wave detectors
    Jade Powell, Daniele Trifirò, **Elena Cuoco**, Ik Siong Heng, Marco Cavaglià, Class.Quant.Grav. 32 (2015) no.21, 215012
- Unsupervised on Real data (ER7):
  - Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data, Jade Powell, Alejandro Torres-Forné, Ryan Lynch, Daniele Trifirò, **Elena Cuoco**, Marco Cavaglià, Ik Siong Heng, José A. Font, Class.Quant.Grav. 34 (2017) no.3, 034002

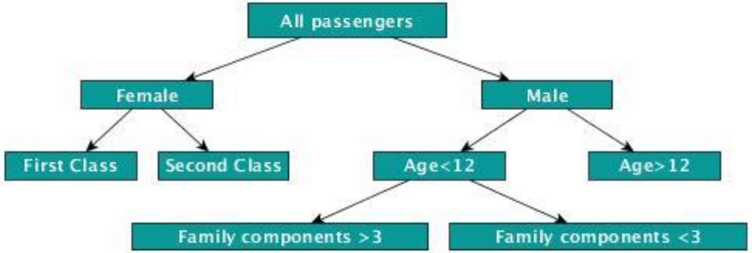


Wavelet Detection Filter and XGBoost (WDFX)



Supervised Classification: eXtreme Gradient Boosting

- https://github.com/dmlc/xgboost
- Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016
- XGBoost originates from research project at University of Washington, see also the Project Page at UW.



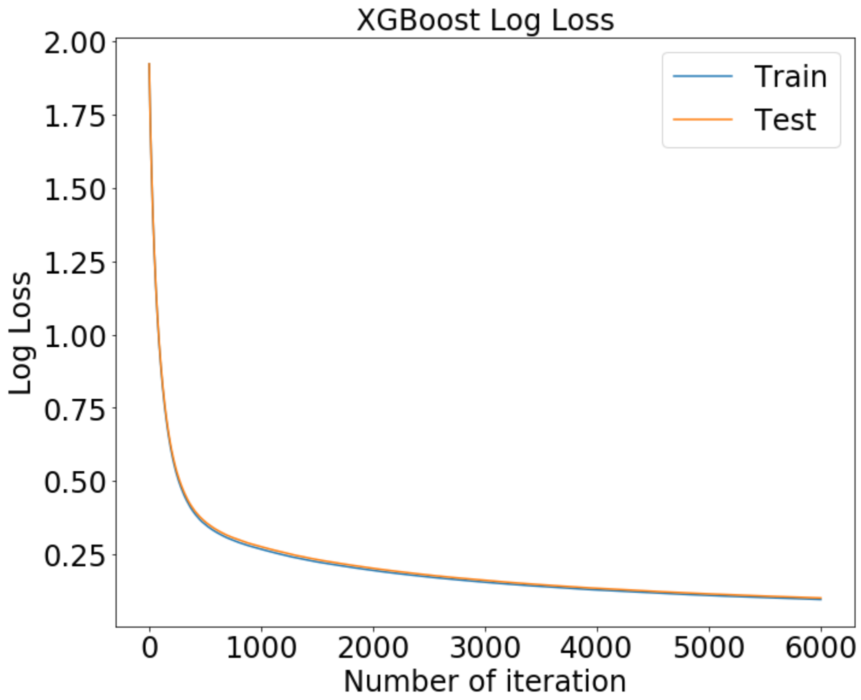
Tree Ensemble

$$y_n = \sum_{k=1}^{K} f_k\,(x_n)$$

dmlc XGBoost

Xgboost

$$L = -\frac{1}{N}\sum_{1}^{N}((y_i \log(p_i) + (1-y_i)(log(1-p_i)) + \Omega$$



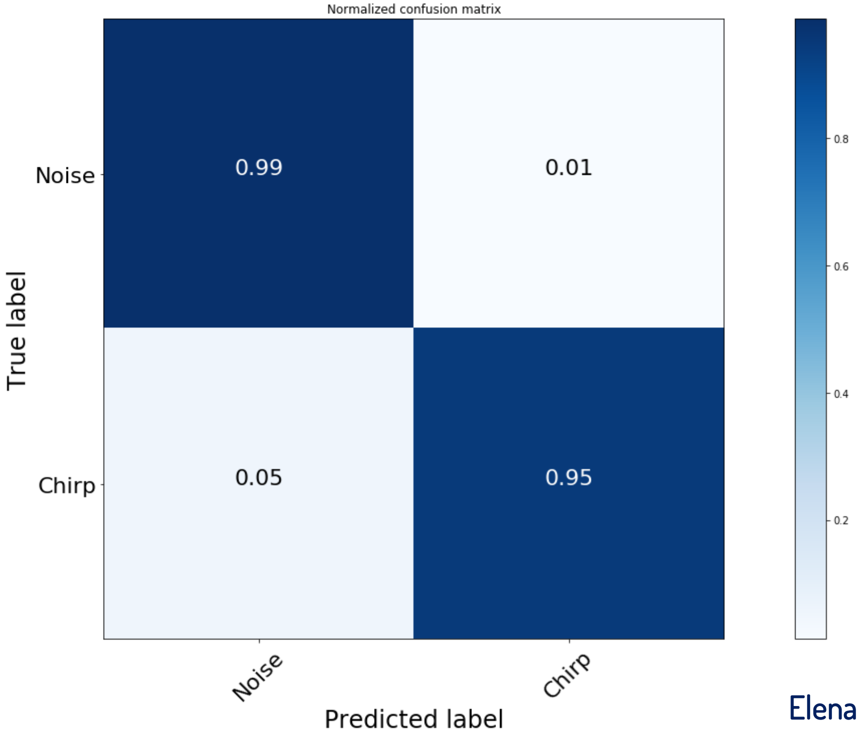
Train/validation/test set: 70/15/15

task	Classes	Learning-rate	Max_depth	estimators
Binary	2	0.01	7	5000
Multi-label	7	0.01	10	6000

WDFX: Binary Classification Results

Overall accuracy >98%

Updated results



Chirp-like signals

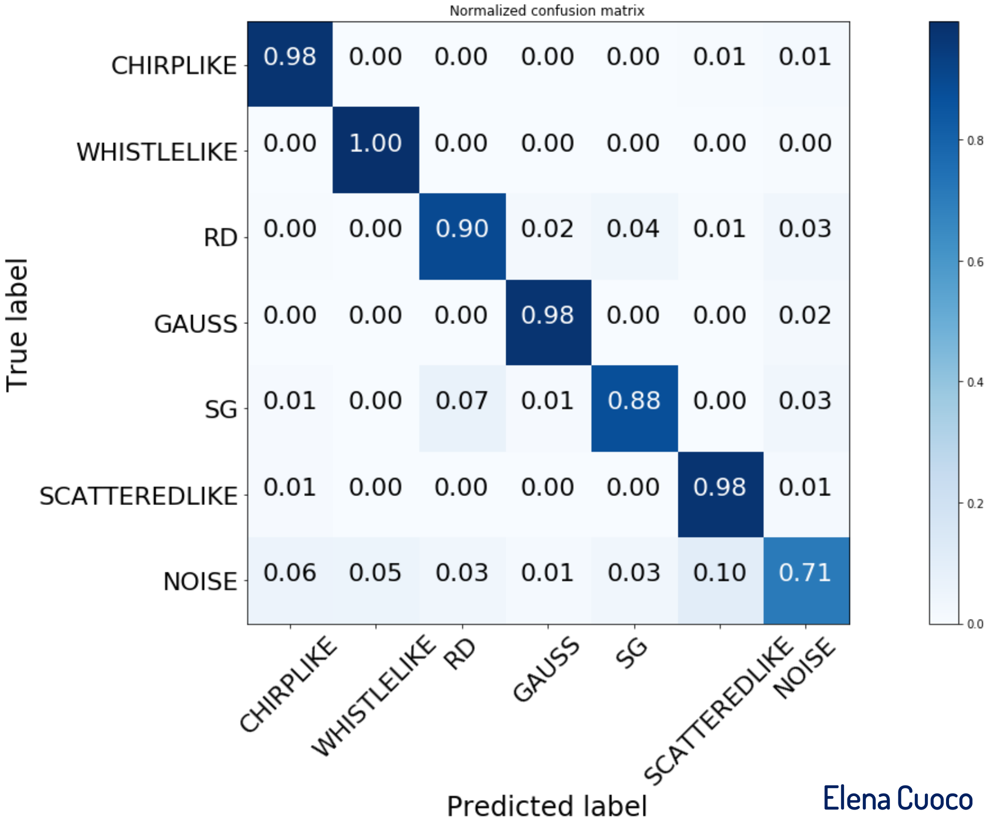
OR

Noise

Cuoco, Razzano in preparation

WDFX Results: Multi-Label Classification

Overall accuracy >93%



Updated results

Cuoco, Razzano in preparation

- Images

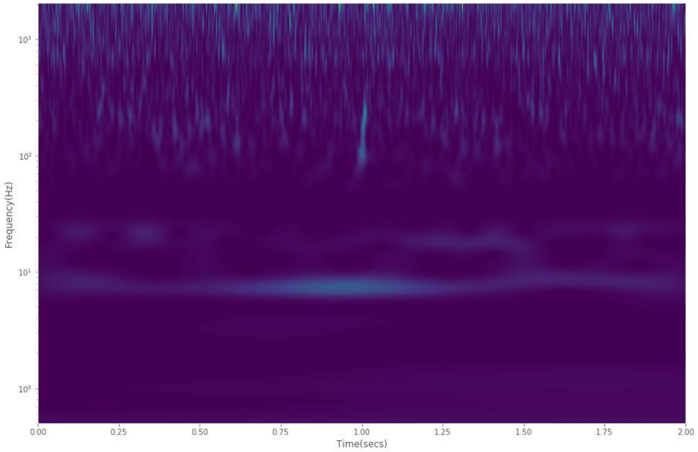
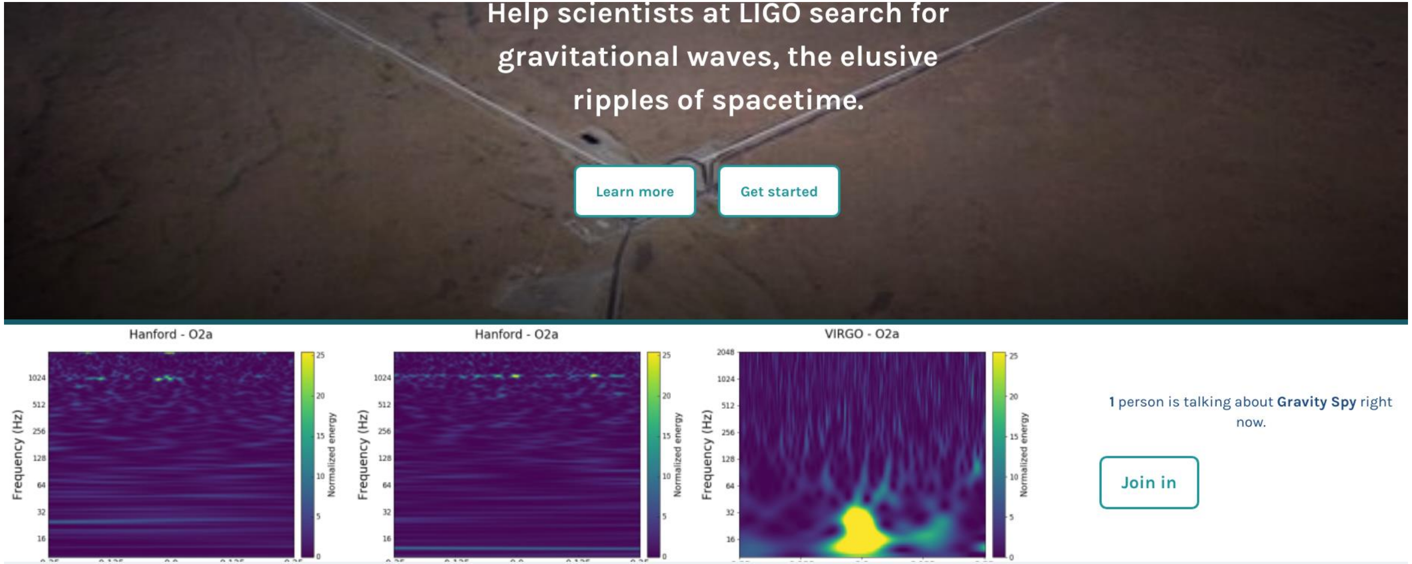


Image-based classification

Glitch & Citizen science: GravitySpy

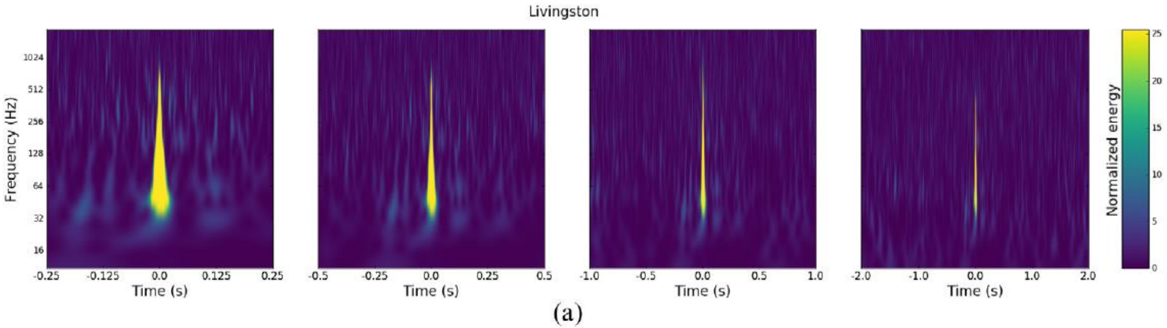


[Www.gravityspy.org](http://www.gravityspy.org)

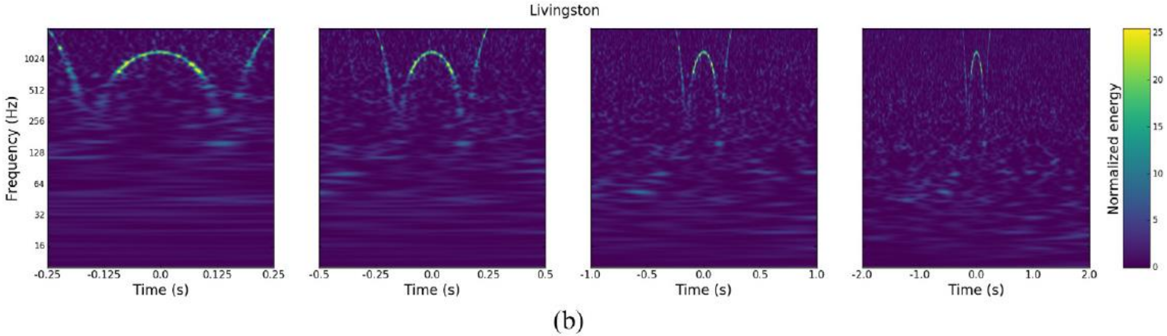
Citizen scientists contribute to classify glitches

More details in Zevin+17

Sample glitch gallery



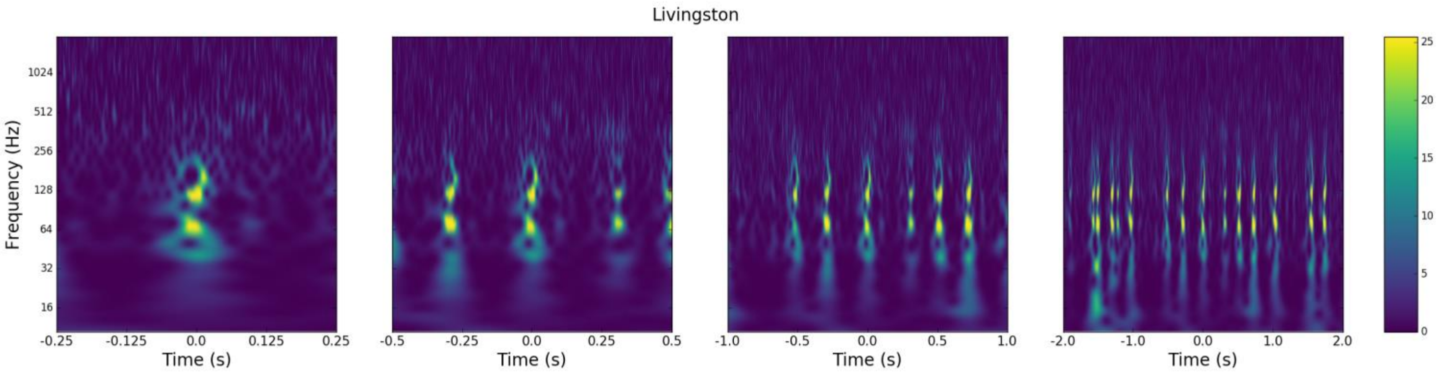
Blip glitches



Whistle glitches

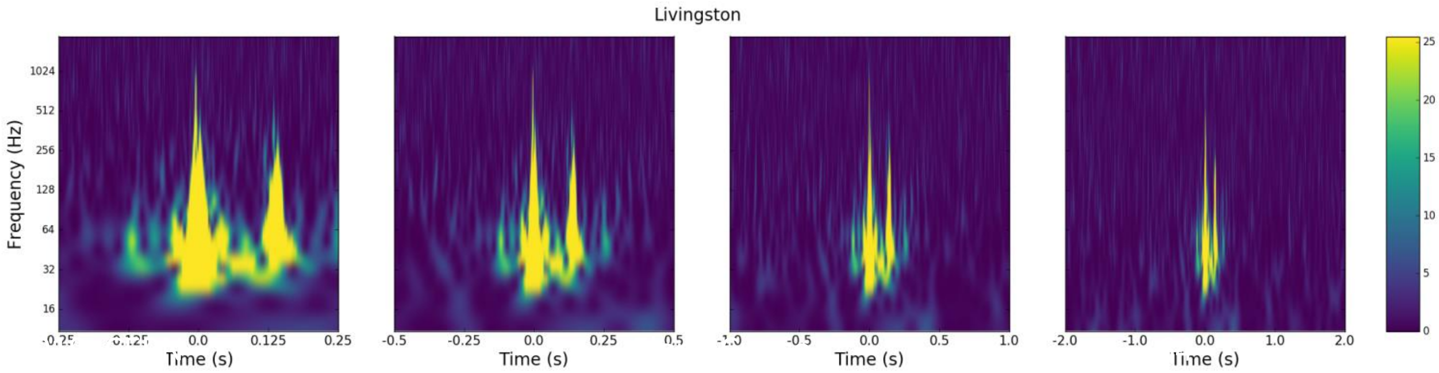
Examples of time-frequency glitch morphology (Zevin+17)

Sample glitch gallery



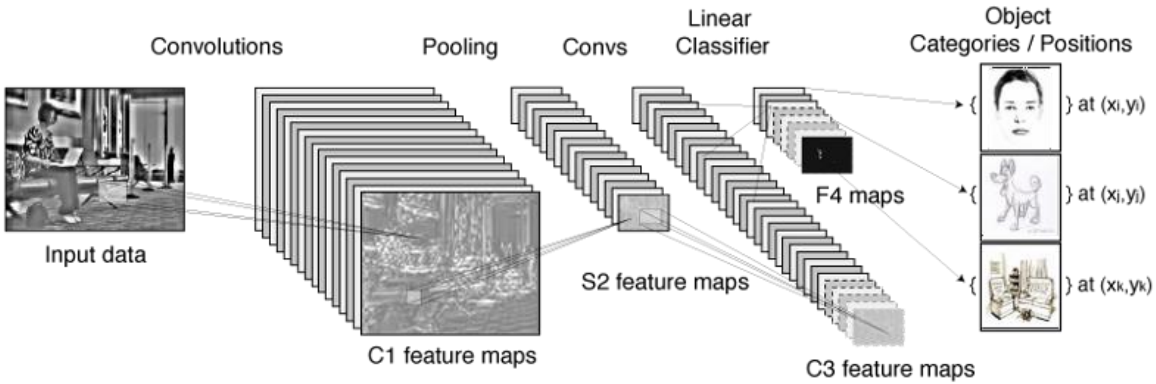
Helix glitches

Koi fish glitches



Deep learning for Glitch Classification

- Many approaches to data: we choose image classification of time frequency images
- The architecture is based on Convolutional deep Neural Networks (CNNs).
- CNNs are more complex than simple NNs but are optimized to catch features in images, so they are the best choice for image classification



Pipeline structure

Input GW data

- Image processing
- Time series whitening
- Image creation from time series (FFT spectrograms)
- Image equalization & contrast enhancement

Classification

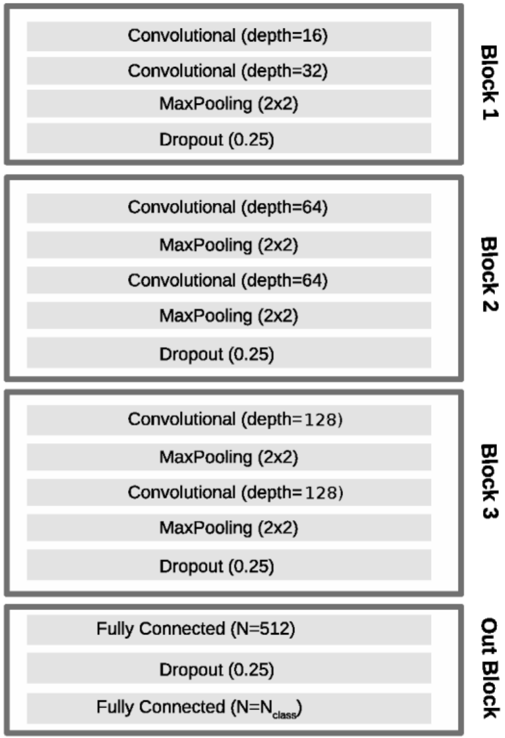
- A probability for each class, take the max
- Add a NOISE class to crosscheck glitch detection

Network layout

- Tested various networks, including a 4-block layers

Run on GPU Nvidia GeForce GTX 780

- 2.8k cores, 3 Gb RAM)
- Developed in Python + CUDA-optimized libraries



Building the images

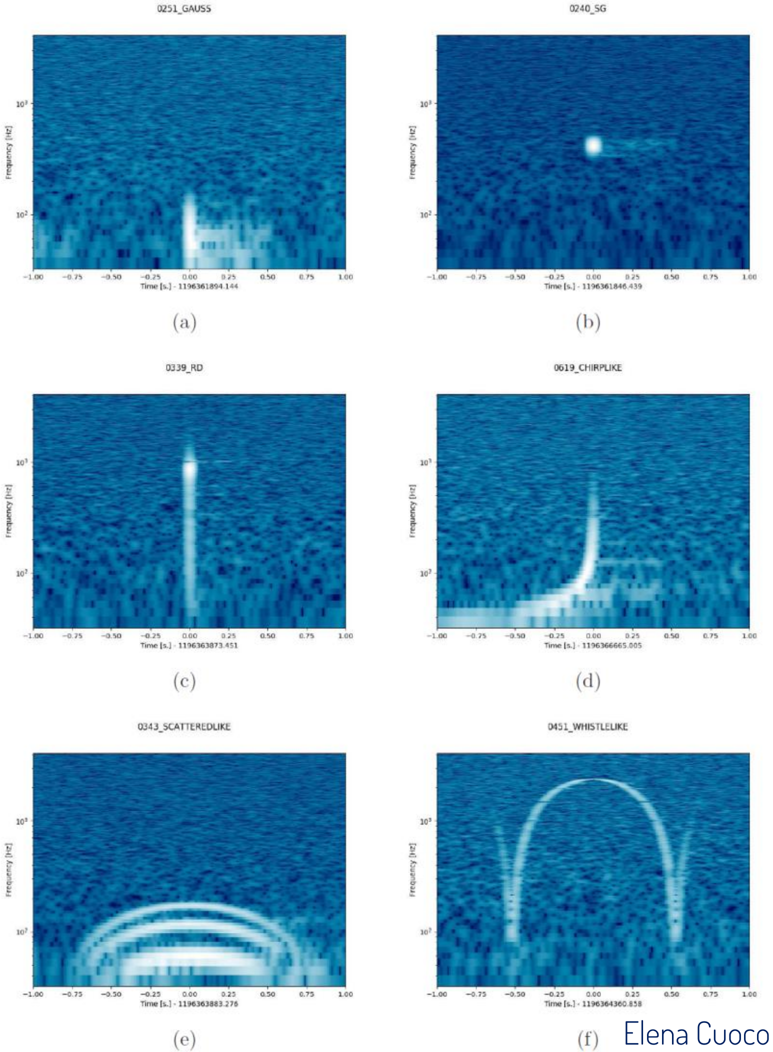
Spectrogram for each image

2-seconds time window to highlight fatures in long glitches

Data is whitened

Optional contrast stretch

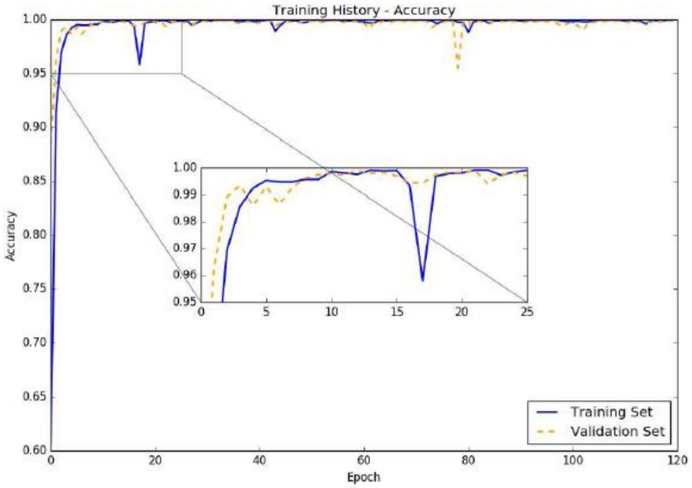
Simulations now available on FigShare

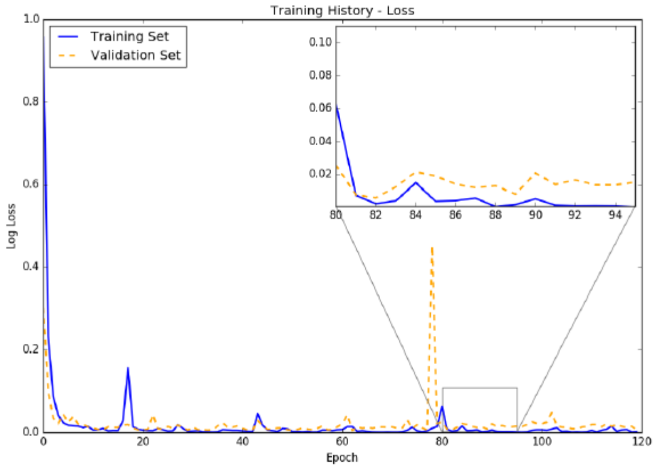


(a) (b) (c) (d) (e) (f)

Training the CNN

- ✓ Datasets of 14000 images
- ✓ Training/validation/test → 70/15/15
- ✓ Image size 241px x 513px
- ✓ Reduced the images by a factor 0.55 due to memory constraints
- ✓ Use validation set to tune hyperparameters
- ✓ On our hardware, training time ~8 hrs for ~100 epochs
- ✓ When training is done, classification requires ~1 ms/image (on our configuration)





Classification Results

We compared classification performances with simpler architectures

Metric	Accuracy	Precision	Recall	F1 score	Log loss
SVM	0.971	0.972	0.971	0.971	0.08
Shallow CNN	0.986	0.986	0.986	0.986	0.04
1 CNN block	0.991	0.991	0.991	0.991	0.02
3 CNN blocks	0.998	0.998	0.998	0.998	0.008

Linear Support Vector Machine

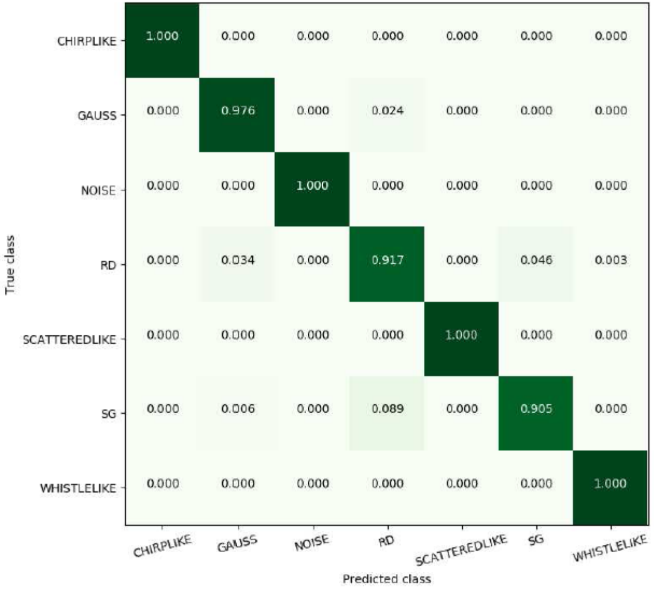
CNN with 1 hidden layer

CNN with one block
(2 CNNs+Pooling&Dropout)

Deep 4-blocks CNNs

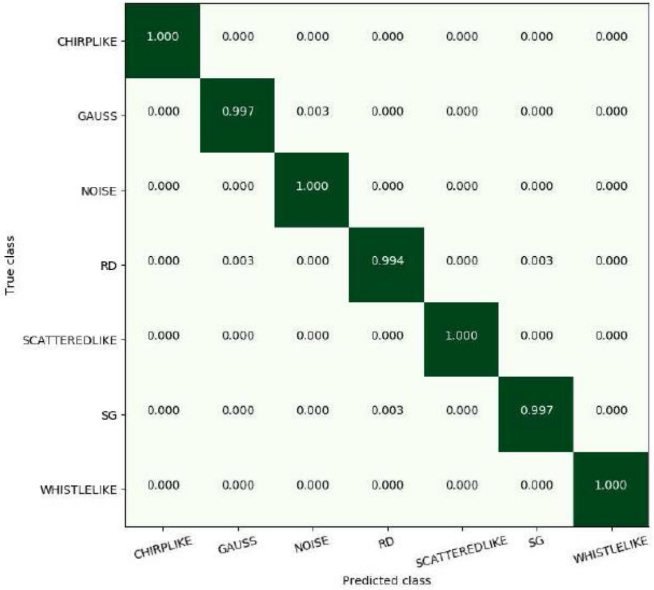
Classification accuracy

Normalized Confusion Matrix



SVM

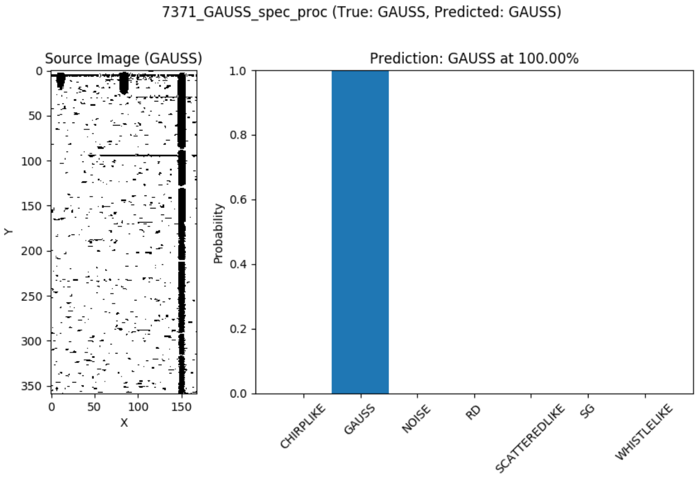
Deep CNN

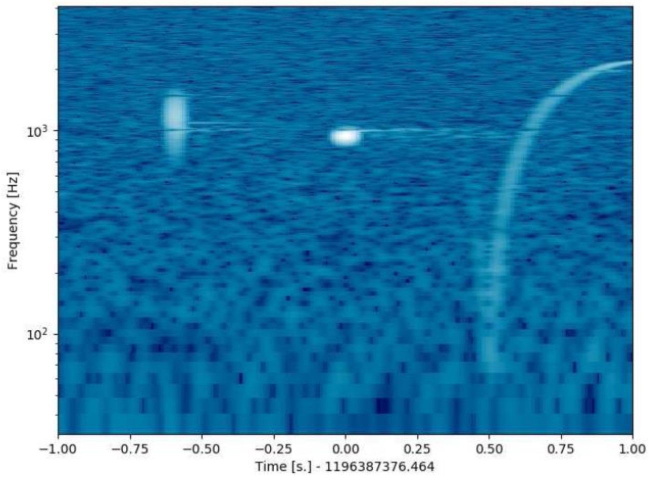


Deep CNN better at distinguishing similar morphologies

Example of classification results

Some cases of more glitches in the time window, always identify the right class





100% Sine-Gaussian

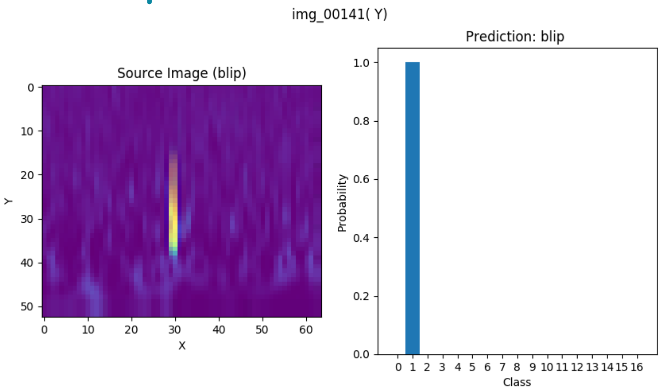
More details in
Razzano & Cuoco 2018, CQG,35,9

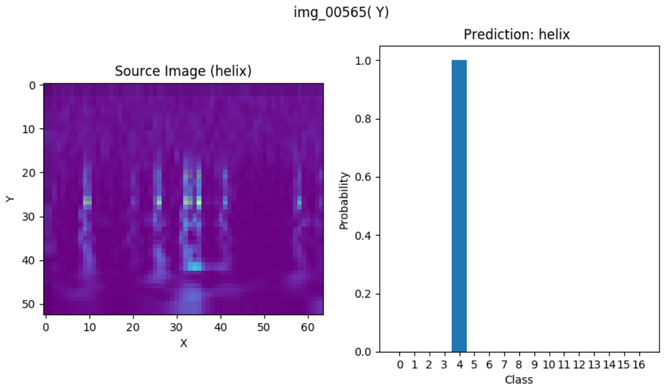
Real data: O1 run

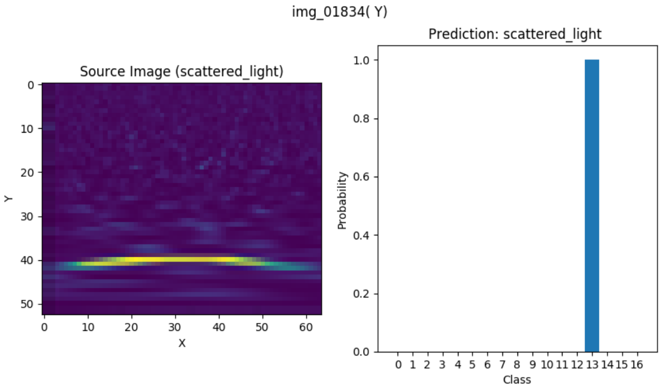
Dataset from GravitySpy images

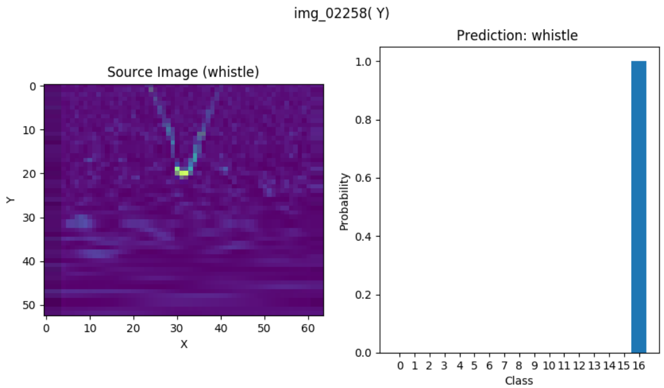
Glitch name	# in H1	# in L1
Air compressor	55	3
Blip	1495	374
Chirp	34	32
Extremely Loud	266	188
Helix	3	276
Koi fish	580	250
Light Modulation	568	5
Low_frequency_burst	184	473
Low_frequency_lines	82	371
No_Glitch	117	64
None_of_the_above	57	31
Paired doves	27	-
Power_line	274	179
Repeating blips	249	36
Scattered_light	393	66
Scratchy	95	259
Tomte	70	46
Violin_mode	179	-
Wandering_line	44	-
Whistle	2	303

Examples of classification

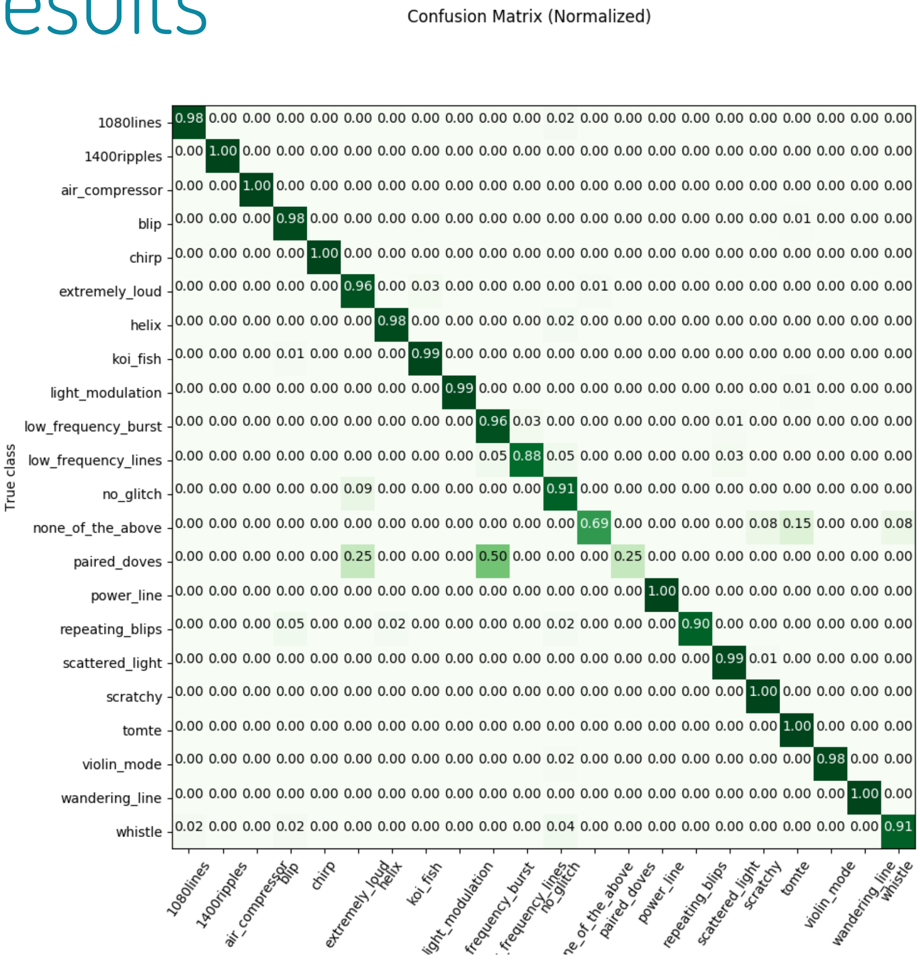








Results



Full CNN stack

Consistent with Zevin+2017

What's next?

- Create a Labelled training set for Virgo data
- Use citizen project to have larger labelled data
- Setup a supervised pipeline running on line on Virgo data
- Use Machine Learning for noise cancellation
- Use Machine Learning for control system

Wavefier

Real time Gravitational Wave transient signal classifier

A project in collaboration with LAPP and Trust-IT services

LAPP Laboratoire d'Annecy de Physique des Particules

cnrs IN2P3 Les deux infinis

EGO EUROPEAN GRAVITATIONAL OBSERVATORY

Trust it Trust-IT Services Communicating ICT to markets

Asterics Astronomy ESFRI & Research Infrastructure Cluster

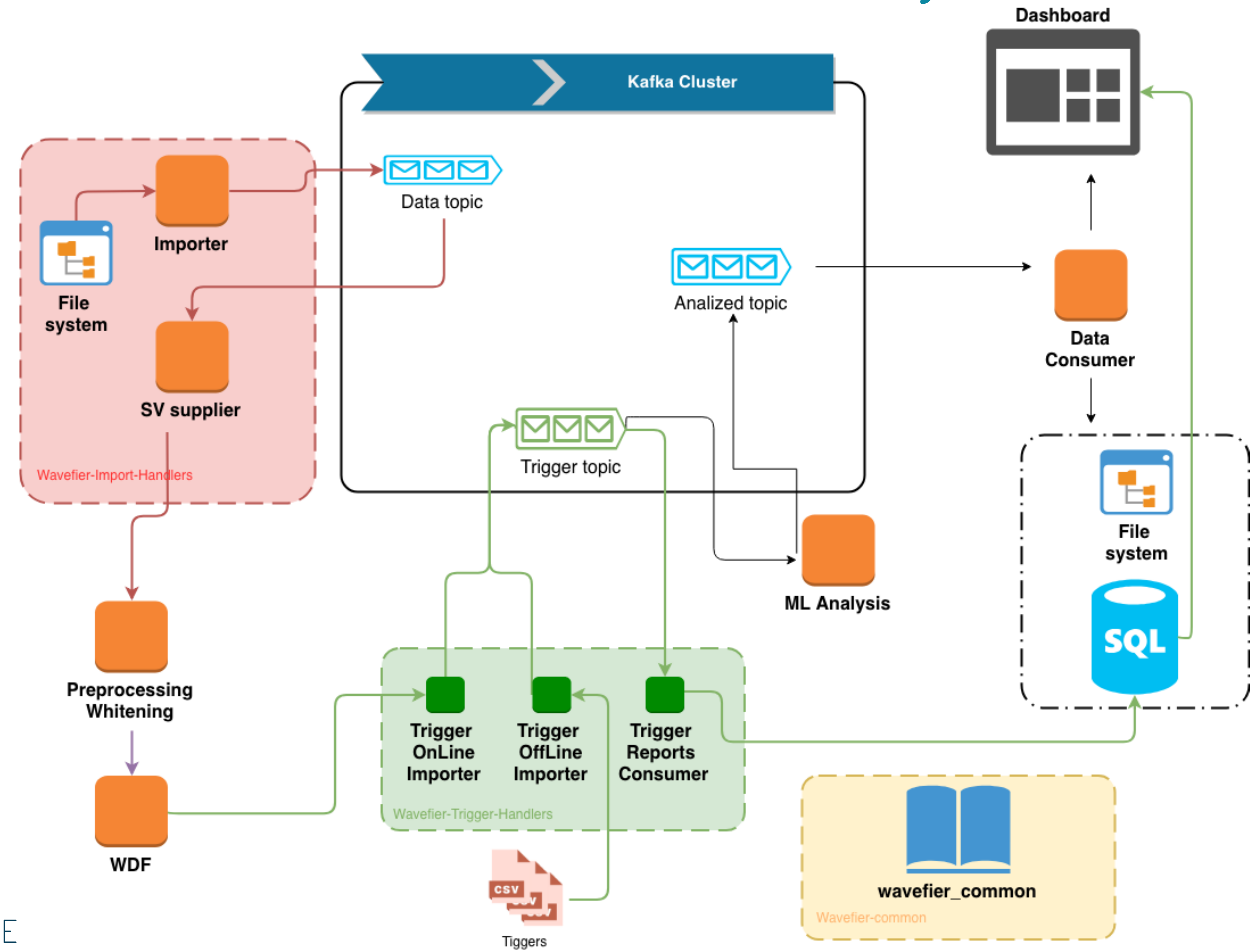
H2020-ASTERICS project brings together for the first time scientists and communities from astronomy, astrophysics, particle astrophysics & big data.
http://www.asterics2020.eu

H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

Wavefier
Real time Gravitational Wave transient signal classifier

Wavefier: real time analysis

Stream to Stream

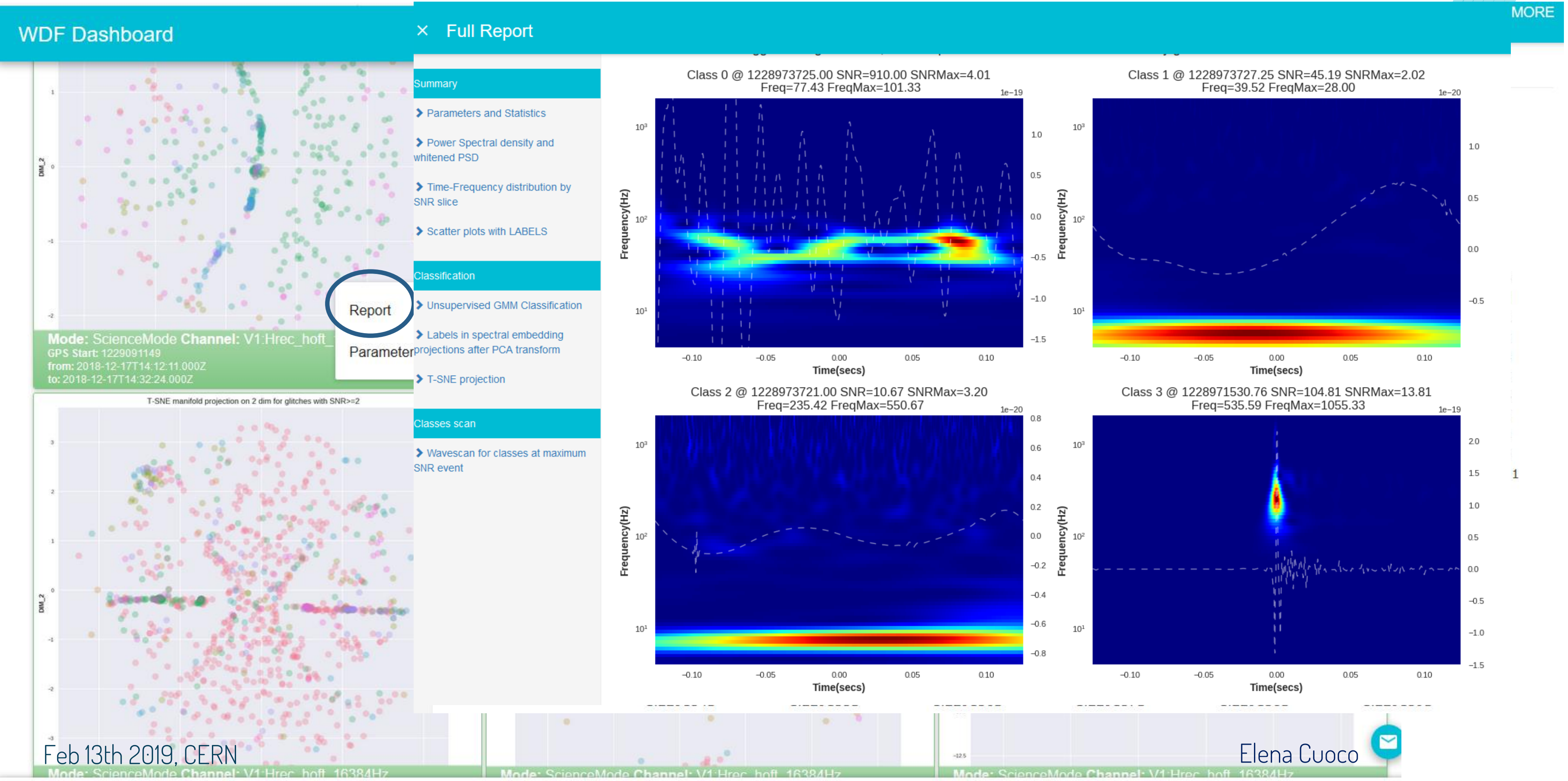


Different Machine Learning approaches

- Wavelet coefficients and some meta-parameters
- Reconstrutcted waveform in 1-D
- Images and CNN
- Transfer learning
- Semi supervised
- GANs to have a larger data set

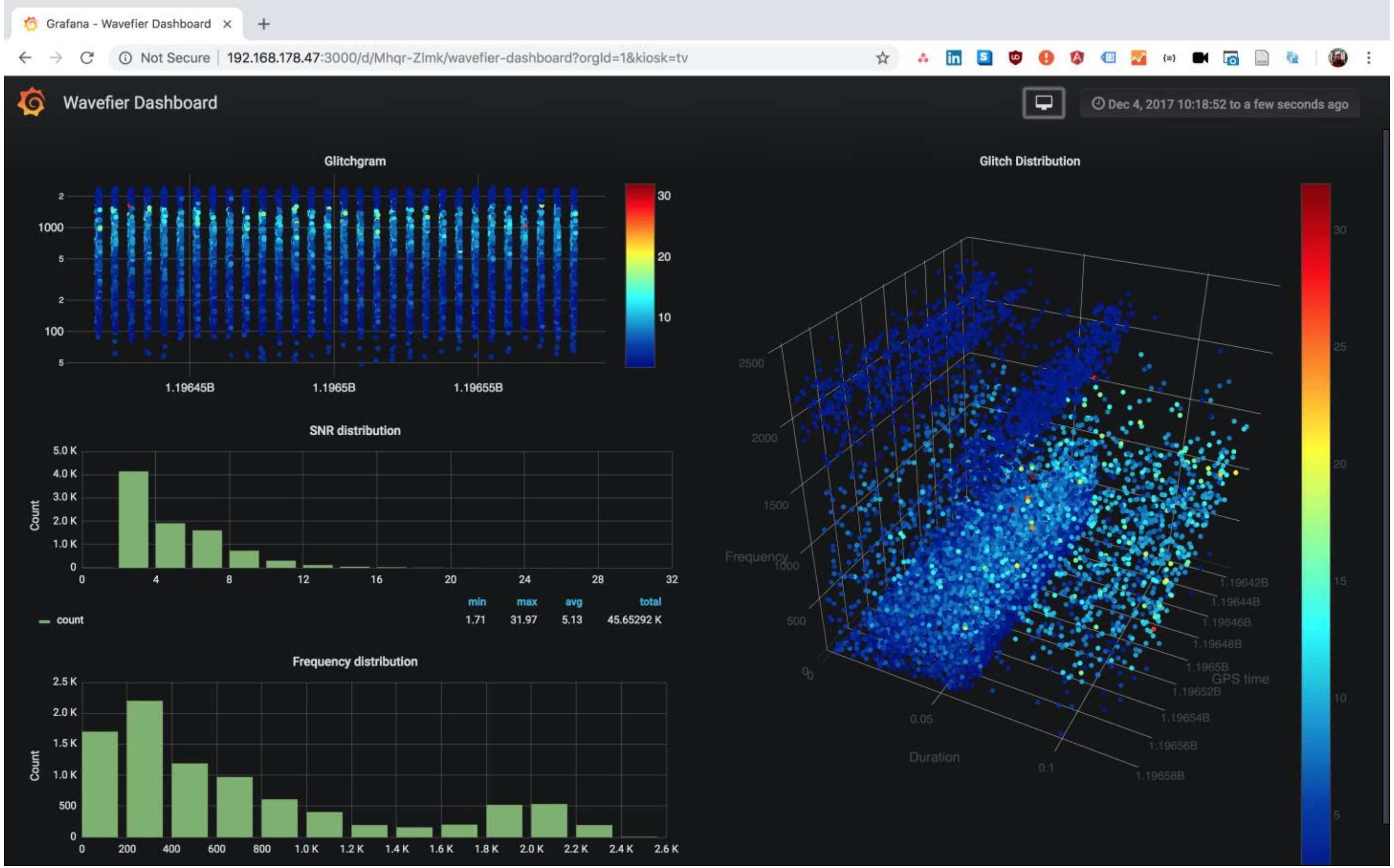
Wavefier
Real time Gravitational Wave transient signal classifier

Showing the results



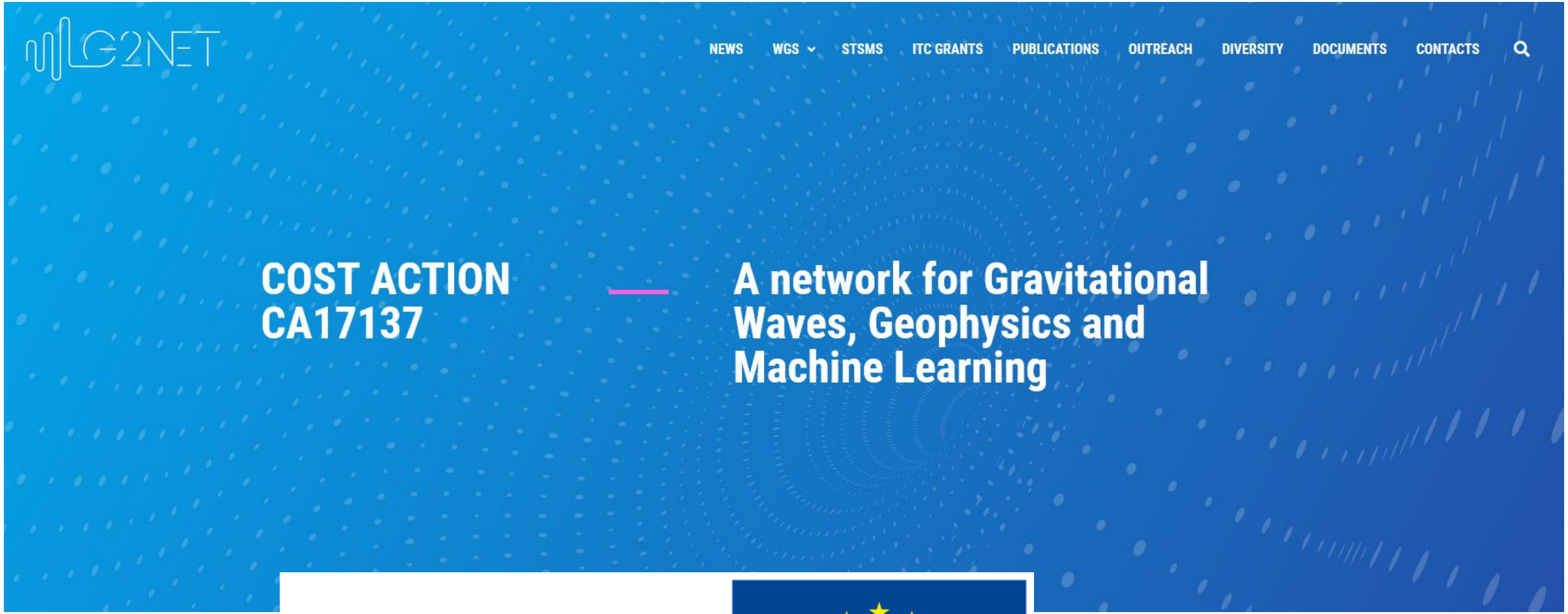
Wavefier Real time Gravitational Wave transient signal classifier

Wavefier dashboard



Grafana. Web based dashboard

http://www.g2net.eu/



EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

G2net: goals of the ACTION

Facilitate conceiving innovative solutions for the analysis of the data of Gravitational Wave (GW) detectors.

Investigate possible solutions to monitor the low-frequency Newtonian noise through the use of adaptive robots.

Train a new generation of young scientists with broad skills in Machine Learning, GW, Control and Robotics.

Investigate new strategies for the handling/suppression of instrumental and environmental noise using Machine Learning techniques.

Bridge the gap between the disciplines of GW physics, geophysics, computer science and robotics

Elena Cuoco

Head of Data Science Office at EGO

SNS Faculty associate

CA17137 g2net Action Chair

ESCAPE General Assembly Chair

You can find me

elena.cuoco@ego-gw.it

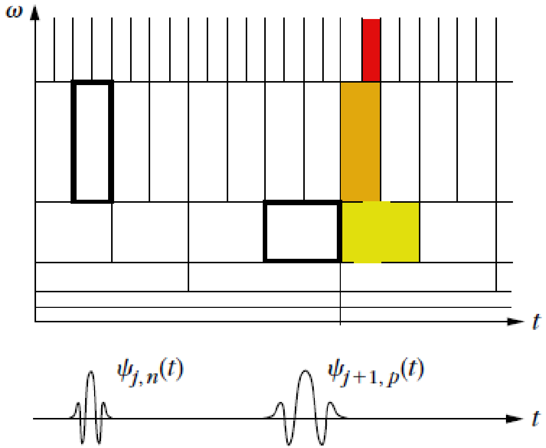
Twitter: @elenacuoco

website: www.elenacuoco.com

THANKS!

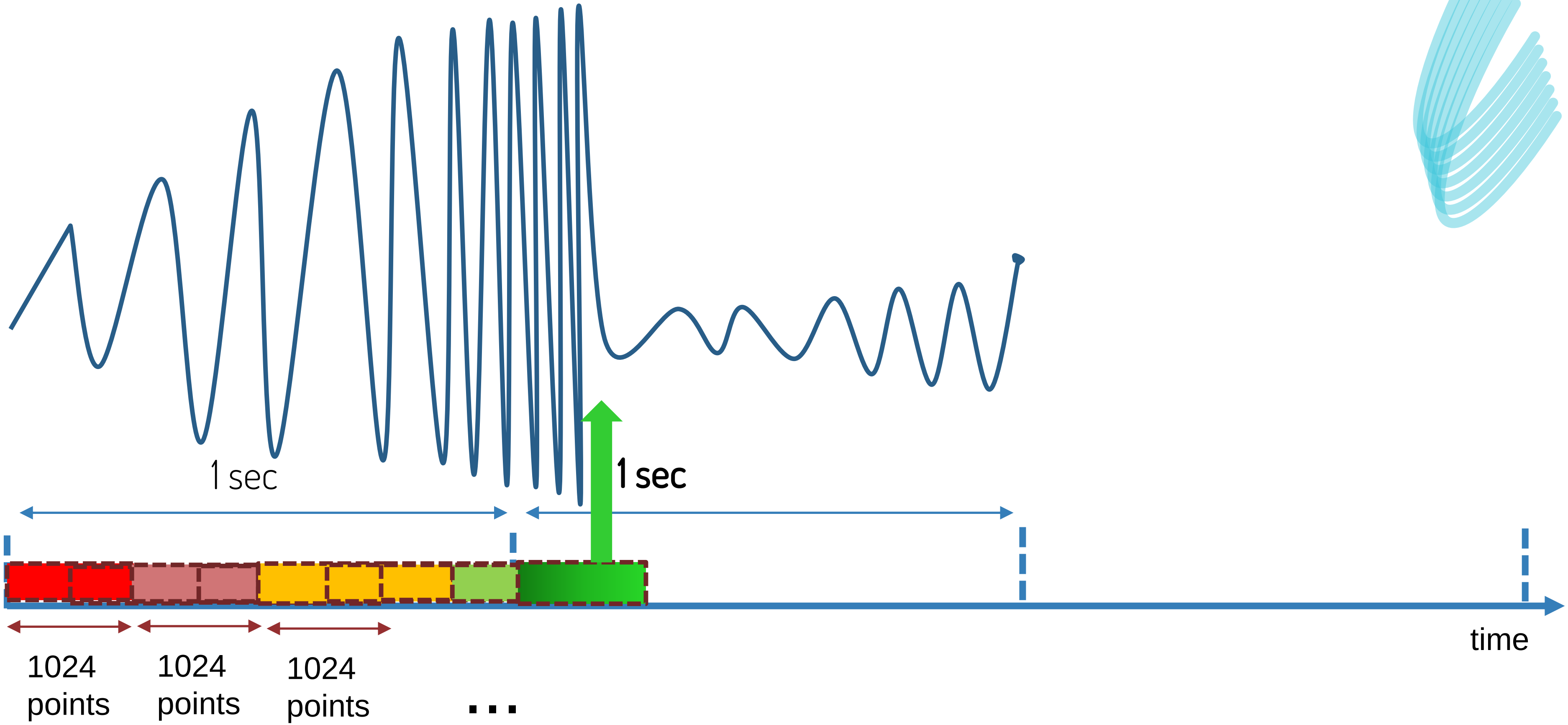
WDF waveform extraction

- Wavelet transform in the selected window size
- Retain only coefficients above a fixed threshod (Donoho-Johnston denoise method)
- Create a metrics for the energy using the selected coefficients and give back the trigger with all the wavelet coefficients.
- In the wavelet plane, select the highest values coefficients to build the event
- Inverse wavelet transform
- Estimate mean and max frequency and snr max of the cleaned event

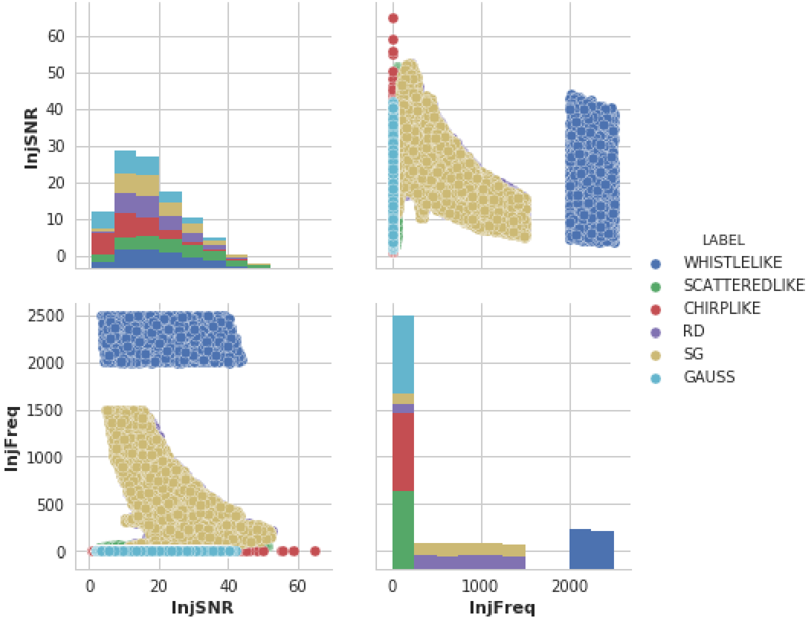


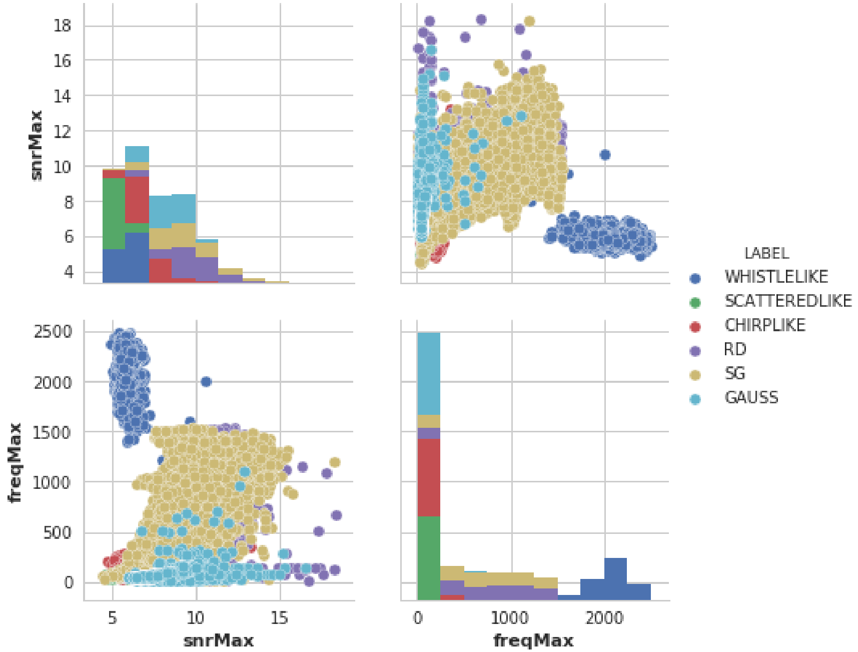
Gps, duration, snr, snr@max, freq_mean, freq@max, wavelet type triggered + corresponding wavelets coefficients.

WDF: how it works



WDF results on simulated data

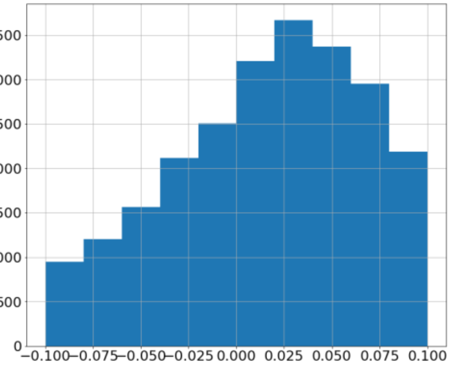




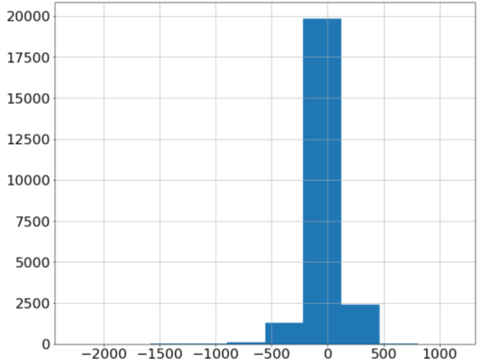
- Detected 99% of injected signals (some with SNR=1)
- False Alarm rate:10% for a time window shift of 1sec for SNR>10

Parameter estimations in 0.1sec

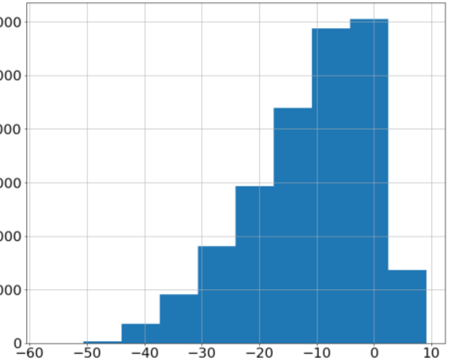
Time diff



Frequency diff



SNR diff



Duration diff