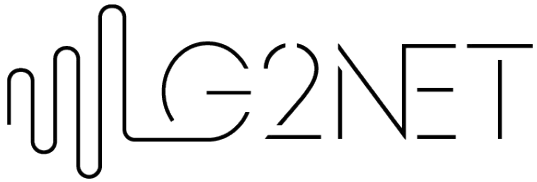
WG1 Introduction

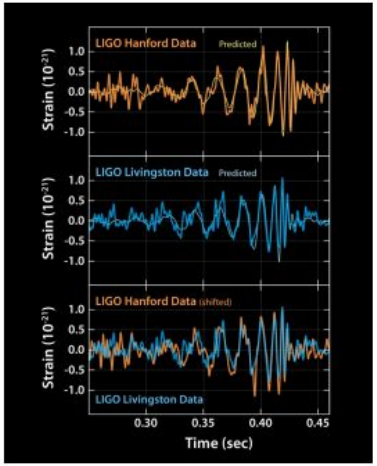
ML for GW astronomy

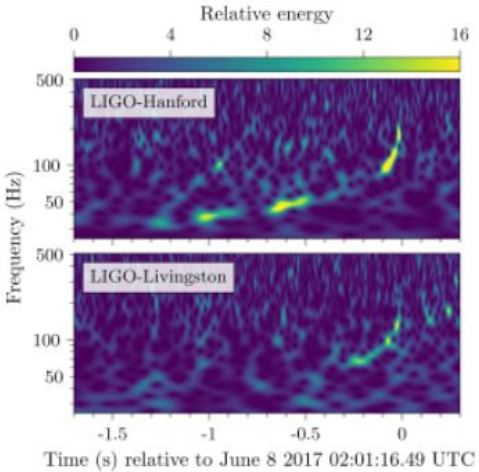
Michał Bejger
Annalisa Appice

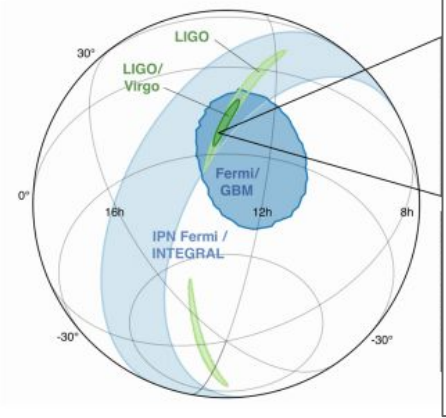
15.01.2019



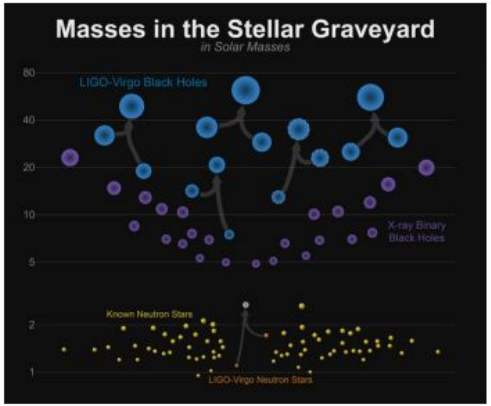
COST ACTION CA17137
A NETWORK FOR GRAVITATIONAL WAVES, GEOPHYSICS AND MACHINE LEARNING

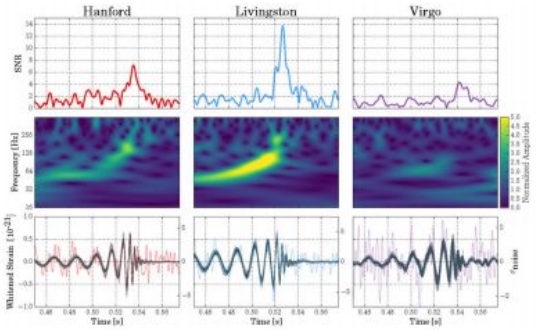


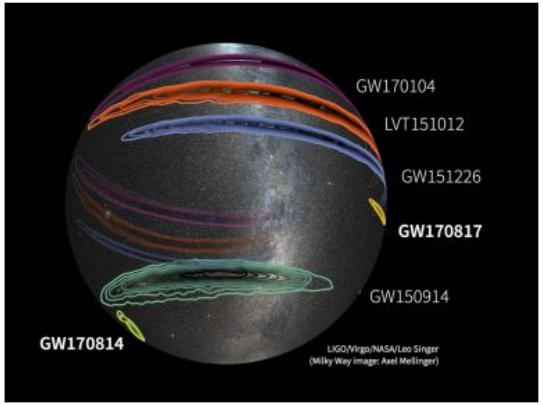




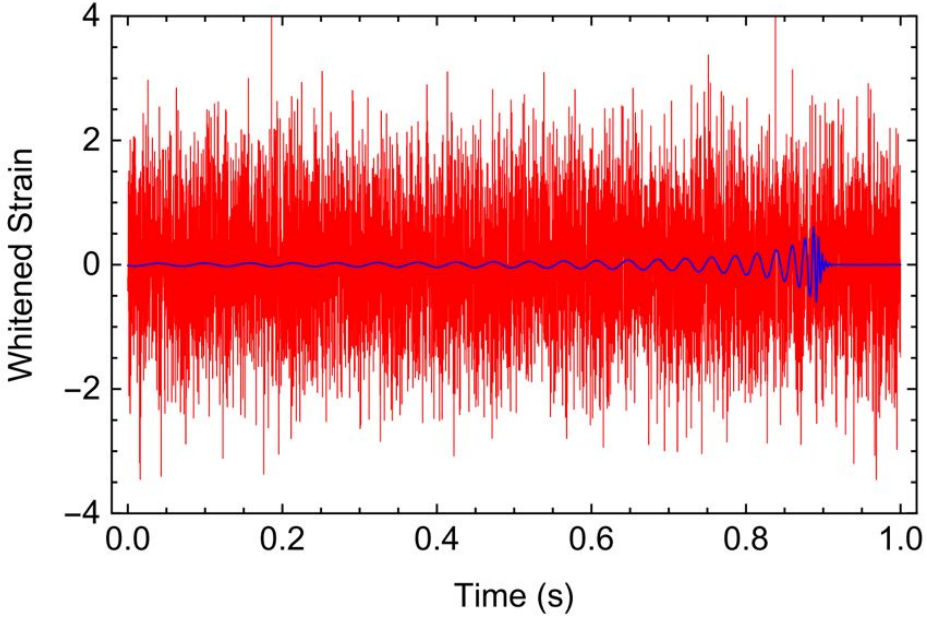
Digging out the signals, classifying them, estimating their parameters **(see Toni's talk)**

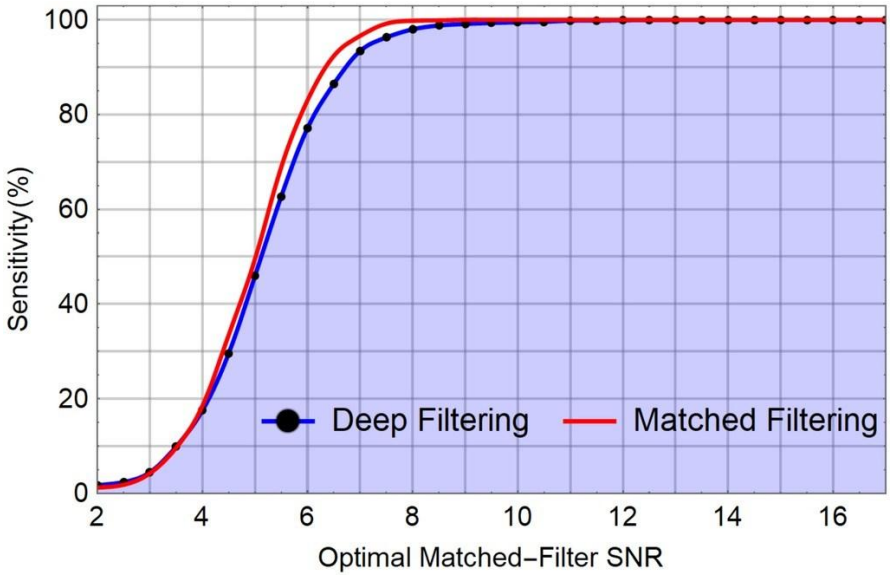






Example: Deep Learning in GW detection





E.g. already existing implementation: George & Huerta ‘’Deep Filtering’’ (based on Convolutional Neural Networks)

First things first: the GW data


LIGO Hanford Observatory, Washington
(Credits: C. Gray)


LIGO Livingston Observatory, Louisiana
(Credits: J. Giaime)


Virgo detector, Italy
(Credits: Virgo Collaboration)

The Gravitational Wave Open Science Center provides data from gravitational-wave observatories, along with access to tutorials and software tools.

Get started!

Download data

GWTC-1: Catalog of Compact Binary Mergers

Join the email list

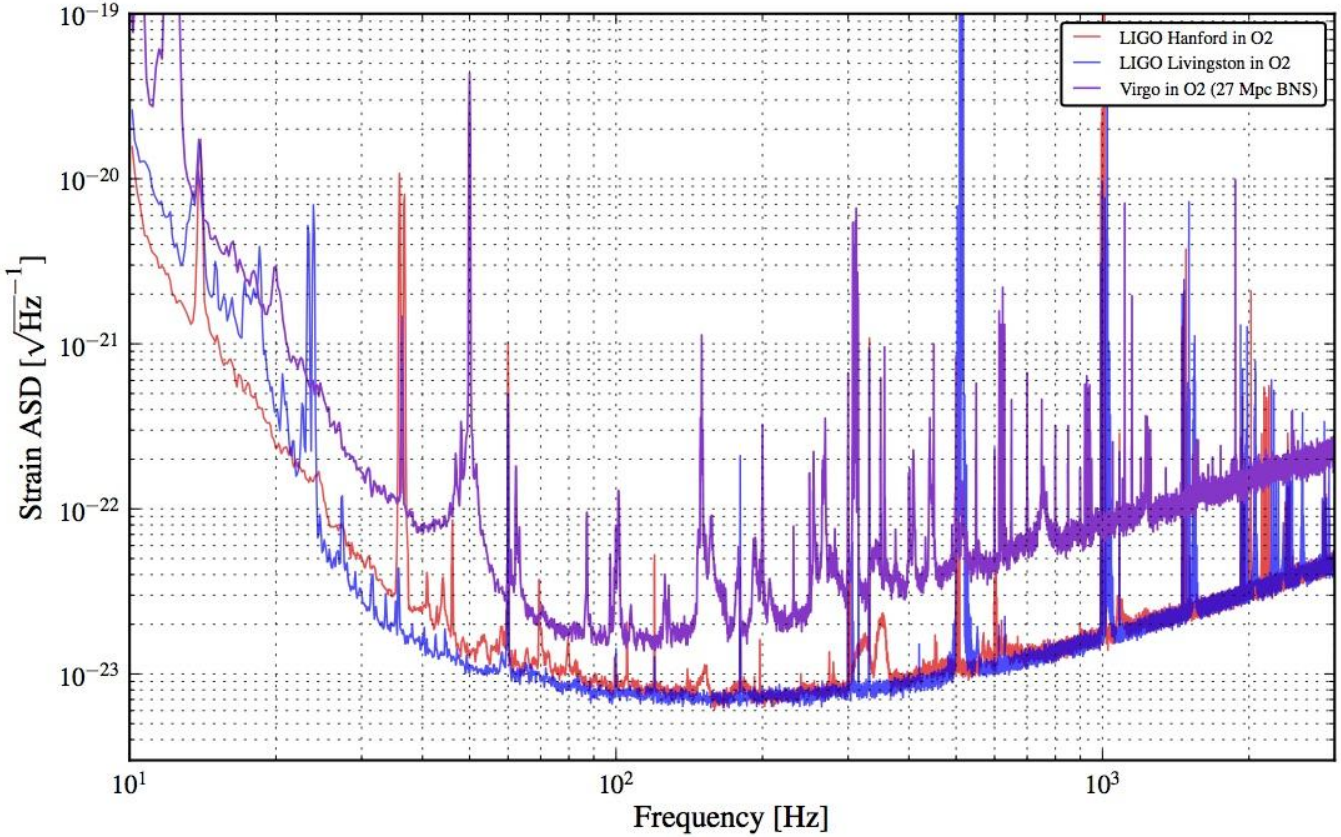
Attend an open data workshop

Data and code resources at the GWOSC (Gravitational Wave Open Science Center)

https://www.gw-openscience.org/about

(see Agata's talk)

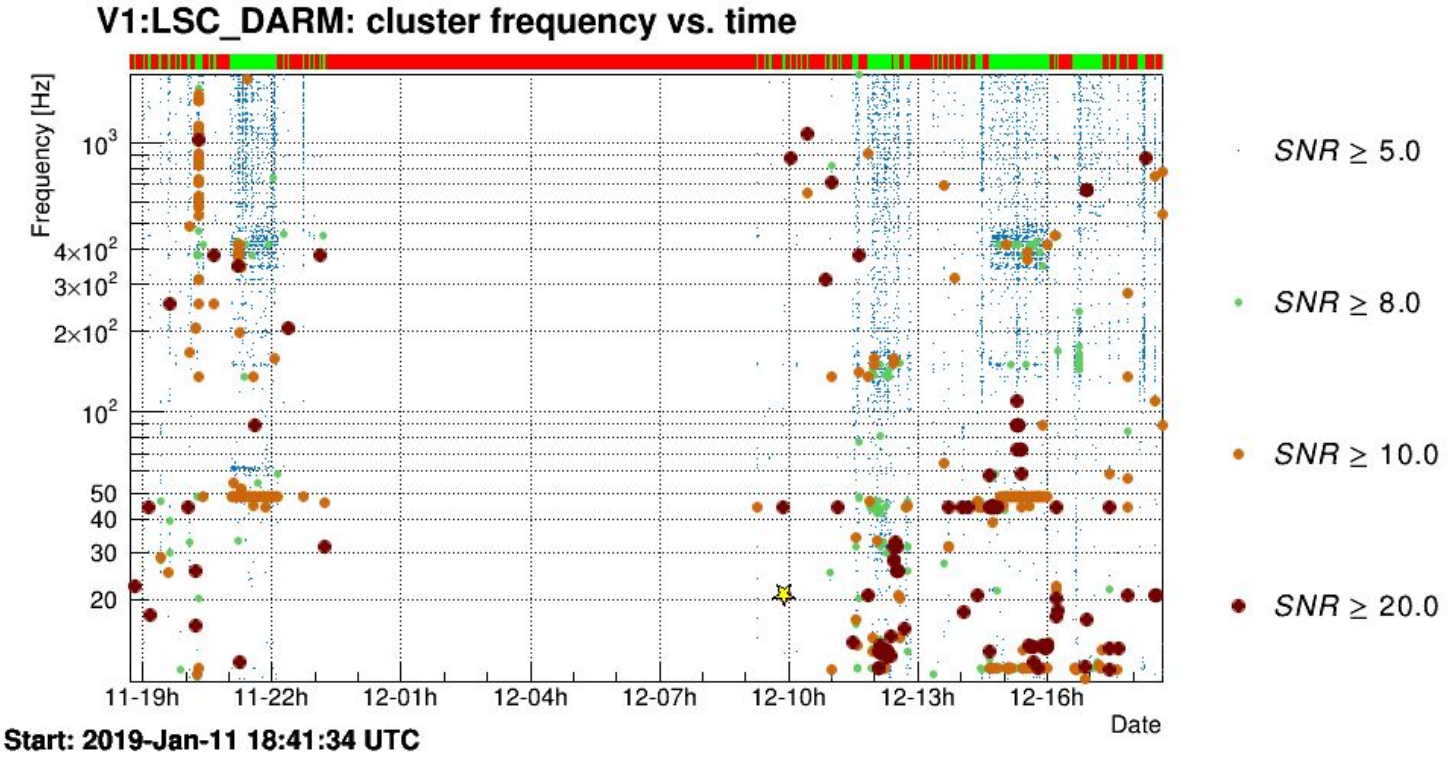
Noise-dominated broadband detectors



- Astrophysical signals are hidden in the detectors' noise
- Many sources of noise: environment, power mains, calibration lines etc. **(talks by Irene, Alessio, Alberto, Gabriele)**

GW-related data analysis

Local disturbances: glitches and spectral features **(see Massimiliano's talk)**



WG1 tasks (from the MoU)

1. Investigate ML classification schemes for GW glitches
2. Design ML pattern-recognition techniques to identify non-stationary spectral lines and non-stationary noise sources
3. Evaluate possible HPC solutions for DL pipelines for online glitch classification
4. Investigate how to classify candidates signals derived from the GW searches
5. **??? (insert your idea here!)**
6. Investigate how to repair the glitched data

WG1 in terms of collaborators

35+ members so far on the mailing list, with an impressive range of skills and interests:

- classical machine learning
- data-driven analysis
- pattern recognition
- signal processing
- time series analysis
- image analysis
- software engineering
- online learning
- big data analytics
- deep learning
- HPC/GPUs cluster computing
- astrophysics
- numerical methods

There is some overlap in goals and skills of other WGs, so we also expect to benefit from these synergies :-)

WG1 practical info

Contact people:

- Michał Bejger (bejger@camk.edu.pl)
- Annalisa Appice (annalisa.appice@uniba.it)

Mailing list: http://mail.ego-gw.it/mailman/listinfo/Wg1-g2net

Teleconferences: ultimately a (bi?)weekly remote meeting **(to be discussed by the group)**

WG1 = ML for GW

- Machine learning (shallow learning and deep learning)
  - Time series analysis and signal Process analysis (classification, clustering, regression, change /anomaly detection, data repair)
  - Image analysis (classification, clustering)
  - Incremental learning to update the patterns as new data are collected
  - Big data architecture to process the big volume of data

Glitch analysis in machine learning: state-of-the-art

- In shallow learning
  - Feature extraction (e.g. Wavelet analysis, Fast Fourier transform + PCA to identify glitch; signal-to-noise ratio, duration, central frequency to characterize glitch)
  - Supervised learning (classification) and unsupervised learning (clustering)

Glitch analysis in machine learning: state-of-the-art

- In deep learning :
  - Time series processed with the sliding window model, artificial glitch injected in the training data
    - Convolution neural networks for detection and/or classification, as well as for regression, in order to estimate the parameters of the source in real-time
    - Long Short-Term Memory (LSTM) neural networks for time series analysis
  - Images derived as spectrograms
    - Convolution neural networks for classification
    - Transfer learning with convolution neural networks (open source weights of CNNs pre-trained on a large dataset of images are used to initialize the CNNs, before fine-tuning (re-training) each model on a training dataset of glitches) ➔ better results with smaller training sets, reduced training time

Glitch analysis vs Big data

- Big data tools to 5V (volume, variety, velocity, veracity, value) dimensions of big data
  - Batch processing (e.g. Apache Hadoop that uses Map/Reduce as a computational paradigm)
  - Stream processing (e.g. Apache Flink - API for unbounded streams in JAVA and SCALA, API for batch data in JAVA, SCALA and Python)
  - Interactive analysis (e.g. JupyteR is an open-source project enabling Big Data analysis, visualization and real-time collaboration on software development across more than a dozen of programming languages.

Glitch analysis vs Big data

(see Roberto's talk)

- The concept of deep learning is to dig large volume of data to automatically identify patterns and extract features from complex unsupervised data without involvement of human, which makes it an important tool for Big Data analysis
  - Deep Belief Network (DBN) that has the capability to learn feature representation from labeled and un-labelled data
  - Convolution Neural Network has many hierarchical layers, consisting of feature maps layers and classification layers. A large number of deep learning methods are connected locally. It is implemented on several hundred cores based on GPU implementation
  - Stacked autoencoder to learn features in unsupervised fashion

Deep Learning in Apache Spark

- It is written in SCALA but it makes available the Python programming model for SPARK, so it will integrate with all of its famous libraries, and right now it uses the power of TensorFlow and Keras, the two main libraries of the moment to do DL.
- It includes high-level APIs for common aspects of deep learning so they can be done efficiently in a few lines of code:
  - Applying pre-trained models as transformers in a Spark ML pipeline
  - Applying Deep Learning models at scale
  - Supporting Transfer Learning
  - Distributed hyperparameter tuning
  - Deploying models in DataFrames and SQL

Additional pieces of machine learning

- Time series analysis
  - Forecasting techniques, Anomaly/Change detection
  - Clustering
  - Interpolation
- Multi-variate regression

Working plan for tomorrow's group meeting

- Define subtasks and identify people responsible for them,
- Identify synergies with other groups,
- Identify necessary data and case studies,
- Sketch actions:
  - milestones,
  - deliverables,
  - collaborations,
  - datasets,
  - new ML tools...

  to be realized in each subtask,

- Discuss organisation of scientific events, STSMs, publications etc.