GWEEP: A Deep Learning Toolkit for Gravitational Waves Analysis

INSTITUTE OF SPACE SCIENCE
A subsidiary of INFLPR

Daniel Tonoiu[1]
Ana Caramete[1]
Laurentiu Caramete[1]
Florentina Pîslan[1,2]

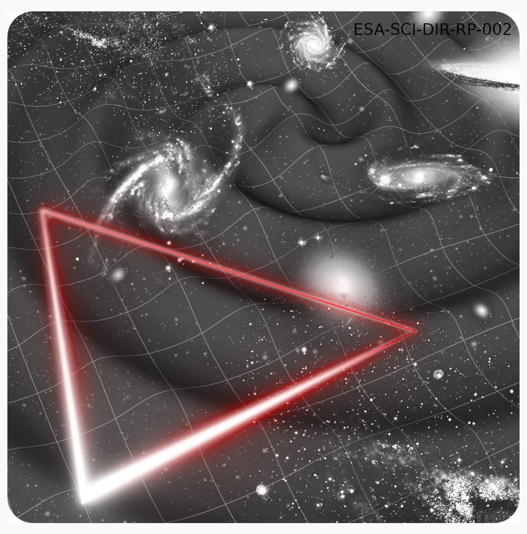
(1) Institute of Space Science, Magurele, Romania
(2) Faculty of Physics, University of Bucharest, Romania

Intro

The Laser Interferometer Space Antenna (LISA) [1] is a selected space mission by the European Space Agency (ESA) aimed at detecting and studying gravitational waves. LISA will provide detailed observations of merging black holes, particularly supermassive black holes, offering insights into galaxy formation and evolution.

LISA will consist of three spacecraft arranged in an equilateral triangle, each separated by 2.5 million kilometers. This triangular formation will act as a giant interferometer, which is essential for detecting gravitational waves.

As demonstrated only a few years ago, during the detection of the GW170817 event [2] by LIGO/Virgo, it is of great importance in a gravitational wave experiment to have a rapid mechanism for alerting other observatories about potential events, so that multi-messenger observations can be performed.

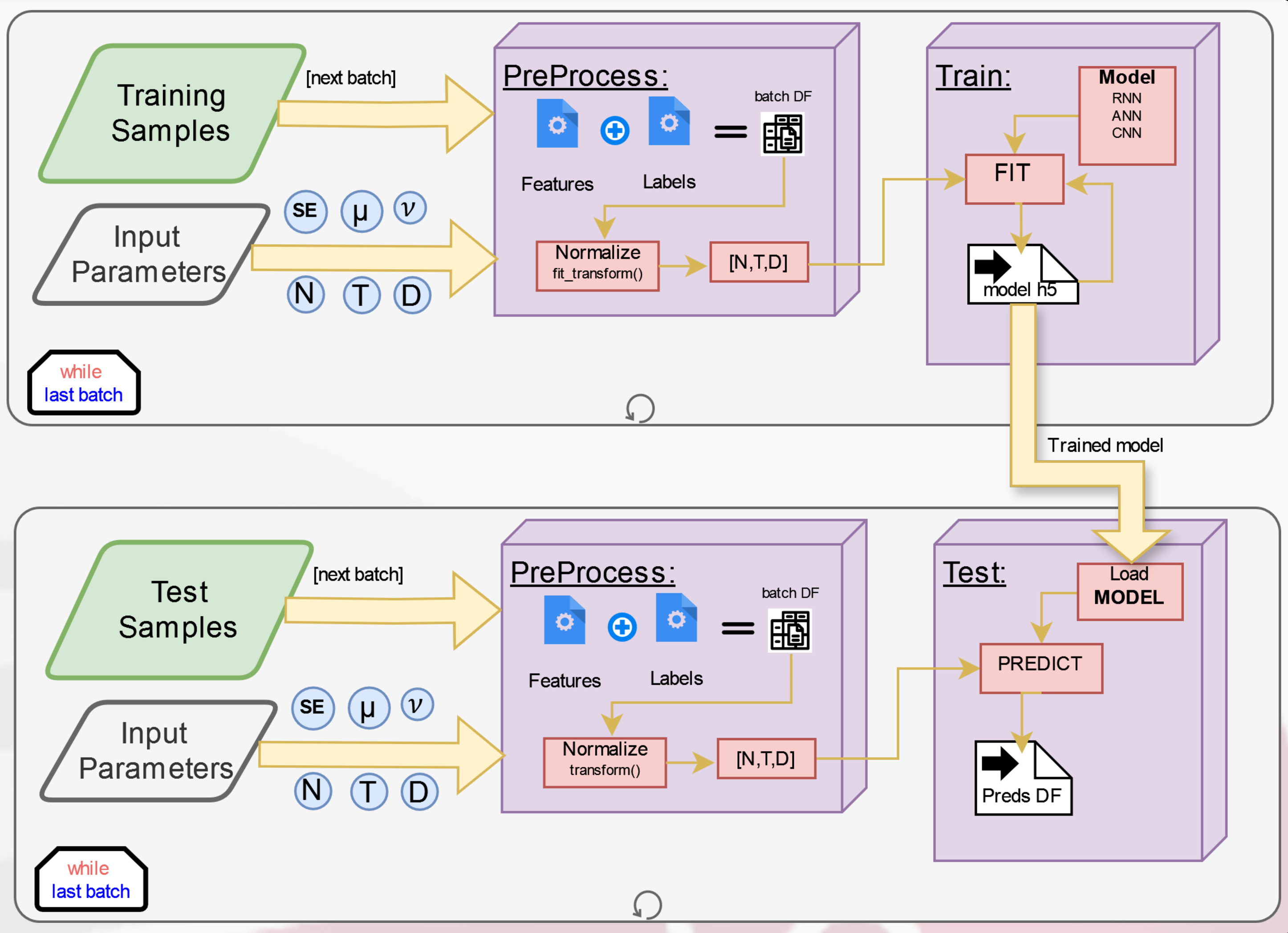


GWEEP Overview

GWEEP is a data analysis toolkit that consists of a collection of neural network processing modules designed to detect and characterize gravitational waves from LISA-like data.

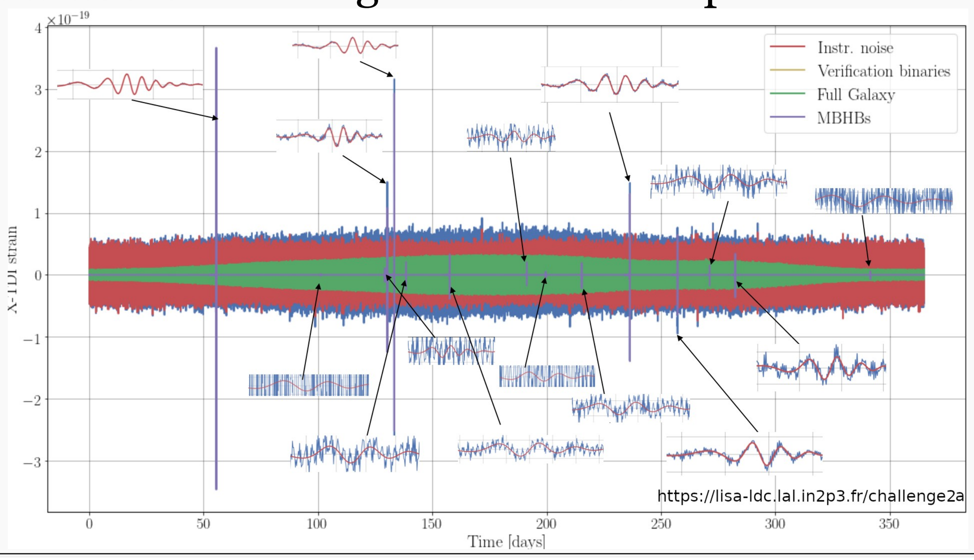
Gweep Components:

- **Preprocess Block**: For each bach of input samples, we calculate the features, generate the labels and normalize the resulting dataframe. We sent a fixed sized [NxTxD] numpy array to the training block
- **Training Block**: The neural network (NN) model is defined and trained, by calling the TF2 *fit()* function. The incremental training technique is used for multiple batches of GW samples. The trained model is saved and then loaded once again with the next batch
- **Test Block**: The test samples are preprocessed, and the trained model from the previous block is loaded. The model predictions are calculated with the *predict()* function and are exported for model evaluation



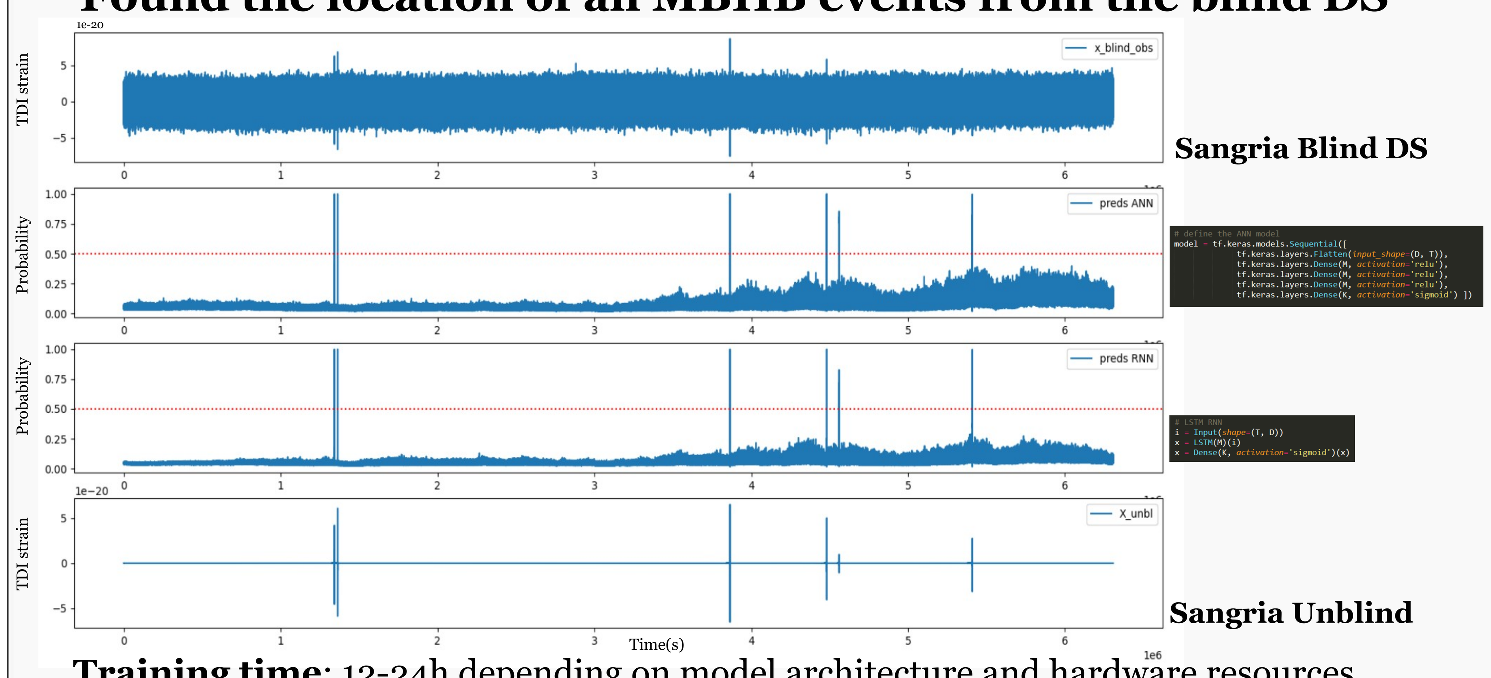
Sangria Data Challenge

- **Objective**: solve the mild source confusion problem [3]
- **Simulated Data**: ~30mln white dwarf binaries mixed with a population of merging massive black hole binaries, with Gaussian noise without gaps
- **Sangria datasets**: Training DS with known parameters + Blind DS



Sangria Results

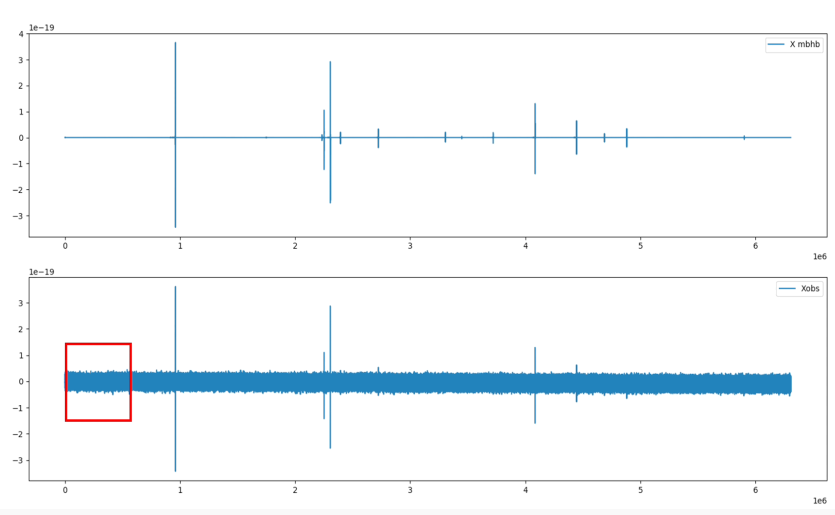
Found the location of all MBHB events from the blind DS



Training time: 12-24h depending on model architecture and hardware resources
Prediction time on Sangria blind: seconds

Methodology

- Sangria training data has only 15 MBHB events, which is not enough to train a NN
- Generate multiple sequences with an **overlapping moving window**
  Labels: every point that belongs to the coalescence interval is labeled as 1 (merger), and otherwise as 0 (inspiral)



- **Training features**:
  - TDI projections on X, Y and Z axes: TDI_X, TDI_Y, TDI_Z
  - TDI noise independent data combinations E and A $E = \frac{X - 2Y + Z}{\sqrt{6}}, \quad A = \frac{Z - X}{\sqrt{2}}$
  - Spectral entropy for each projection $H(x, sf) = -\sum_{f=0}^{f_s/2} P(f)\log_2[P(f)]$
- **Preprocess**:
  - Split the Sangria training data into a train DS of size 68% and a test DS of size 32%. The test data includes the last 4 MBHB events of the Sangria training data
  - Build extended train and test datasets, using an overlapping moving window of size T=10 to generate multiple sequences. This allows the neural network to learn to classify sequences as either part of a merger or an inspiral event
- **Train**:
  - Train on multiple neural network architectures including LSTM, BiLSTM, ANN, CNN, GRU, and SimpleRNN
  - Tune multiple hyperparameters: number of features (D), window size (T), number of hidden units (M), learning rate, and number of hidden layers
- **Test**:
  - Calculate predictions on test data for the last 4 MBHB events
  - Train initially with the simplest NN architecture and gradually increase the model complexity
  - Plot the predictions probability distribution along with the labels

Conclusions and Future Work

- Detection of the MBHB peaks from the Sangria blind data set
- Demonstrate the feasibility of developing a low-latency pipeline for MBHB events
- Future tool development is needed for an increased detection accuracy
- We now work on 1 year worth of data made from 1500 LISA like simulated samples with multi-class classifier for redshift and M1/M2 parameter estimation

References:

1. Danzmann K. et al., Laser Interferometer Space Antenna, Submitted to ESA on January 13th in response to the call for missions for the L3 slot in the Cosmic Vision Programme, arXiv:1702.00786, 2017
2. B. P. Abbott et al., "Properties of the Binary Neutron Star Merger GW170817", 2019, Phys. Rev. X 9, 011001, DOI: 10.1103/PhysRevX.9.011001
3. Quentin Baghi, The LISA Data Challenges, arXiv:2204.12142, 2022

Acronyms:

- **ANN**: Artificial Neural Network
- **CNN**: Convolutional Neural Network
- **RNN**: Recurrent Neural Network
- **DF**: DataFrame
- **DS**: DataSet
- **GWEEP**: Gravitational Wave DEEp Learning Pipeline
- **GW**: Gravitational Waves
- **NxTxD**: no of samples x sequence size x no. of features
- **MBHB**: Massive Black Hole Binary
- **SE**: Spectral Entropy
- **TDI**: Time Delay Interferometry
- **TF2**: TensorFlow 2

Supported by the Romanian Ministry of Research, Innovation and Digitalization under the Romanian National Core Program LAPLAS VII - contract no. 30N/2023 and by ESA PRODEX project DPLISAM