# Learning to detect continuous gravitational waves

**Rodrigo Tenorio** Michael J. Williams Chris Messenger

29th September 2022

Machine Learning in GW search: G2net next challenges

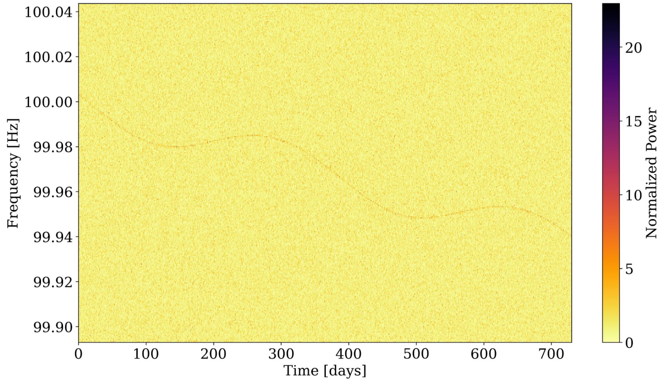
Universitat de les Illes Balears

University of Glasgow

# Introducing continuous gravitational waves (CWs)

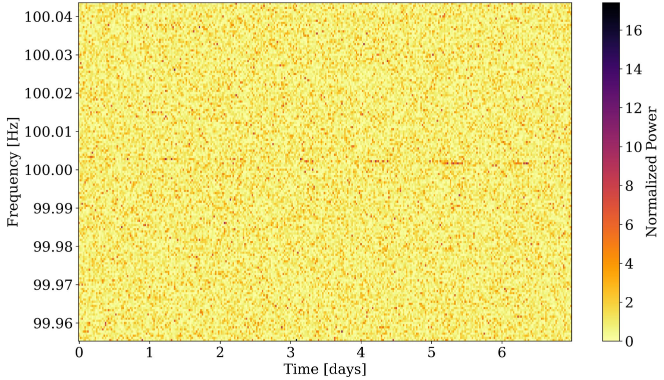
Continuous gravitational waves (CWs):

- Long-standing gravitational waves (months - years).
- No detection to date!
- Expected sources: rapidly-spinning neutron stars (NS).
  - Lots of interesting physics to learn from them.
  - Probe the EM-quiet population of NS.




- Typical CW amplitudes: 2-3 orders of magnitude smaller than that of a CBC signal.
  - Peak strain of GW150914: O($10^{-21}$)
  - CW amplitude upper limit on O3: O($10^{-25}$)

- Simple signal model:
  - Amplitude is constant for the duration of the run
  - Quasi-monochromatic signal.
  - Doppler modulation induced by the Earth's motion.

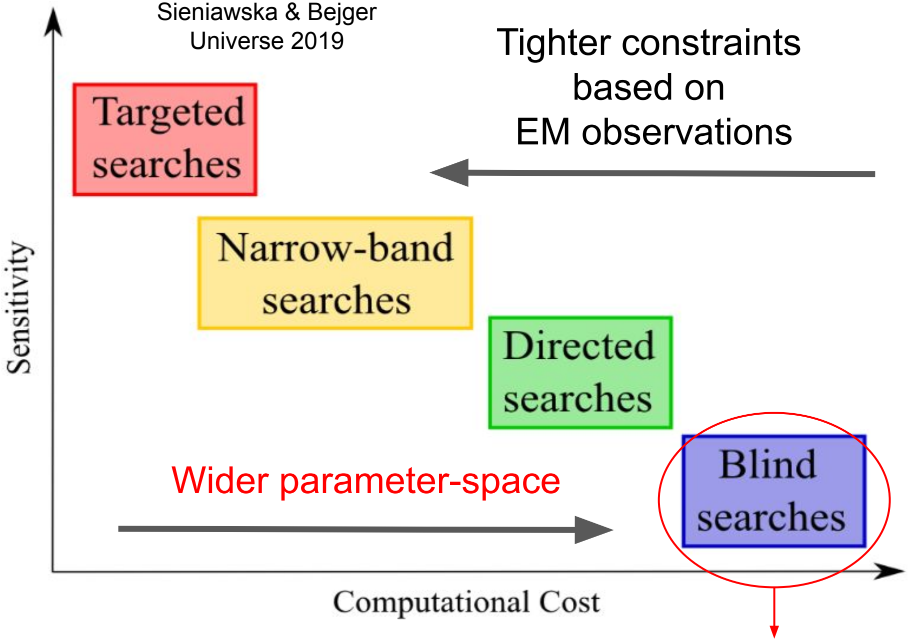


# Searching for CWs

CW searches are classified according to what information goes into them:

- Targeted:
  Focus on a specific pulsar, phase-lock GW radiation to EM emission.
- Narrow-band / directed:
  Focus on a specific sky position, GW emission independent from EM.
- Blind:
  Look for signals from unknown sources with a specific Doppler modulation.

Most non-targeted searches are source agnostic: no specific property is required from a source insofar they emit a quasi-monochromatic signal.




We'll focus on these ones!

First trade-off:

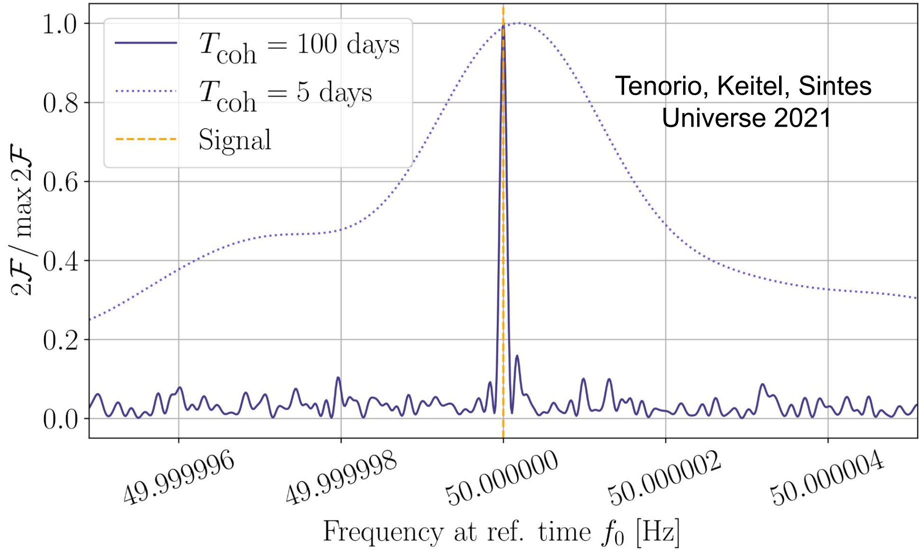
- The less we know about our sources, the wider the parameter-space to cover.
- Relying on EM observations may drive us away from the GW-loud NS population.

# Searching for CWs

CW signals are significantly longer than CBC signals (months – years vs. seconds).

Matched-filtering (standard tool) is unfeasible for wide parameter-space searches.

- Long signal → Huge template banks.




"Solution": Hierarchical approach

- Use sub-optimal methods to sweep wide parameter-space regions.
  - Impose a looser signal model, reducing the required parameter-space resolution.
- Follow-up interesting candidates with more sensitive methods.

Second trade-off:

- Non-optimal methods are less sensitive than matched-filtering for a given model.
- Non-optimal methods perform better than matched-filtering at a fixed computing cost.

# State of the art of CW searches

Less candidates

Lower computing cost

Most searches are structured as follows:

**Main search stage**

Evaluate a detection statistic over the parameter-space. Accounts for most of the computing cost!

**Post-processing**

Reduce the number of candidates to take care of using some form of clustering.

**Application of vetoes**

Discard obvious candidates produced by instrumental artifacts.

**Follow-up**

Re-analyze surviving candidates using a more sensitive method.

Typical searches run in a few million CPU hours on a supercomputer [Abbott+ PRD 100, 024004 (2019)].

GPU acceleration allows for deeper searches [Wette+ PRD 103, 083020 (2021)] and more diverse signals [Covas & Sintes PRD 99, 124019 (2019), Keitel & Ashton CQG. 35 (2018), 205003].

Broadest and deepest searches to date run on the volunteer-distributed computing platform Einstein@Home, consuming ~100 M CPU hours [Steltner+ ApJ 909 79 (2021)].

Extensive manual tuning, however, tends to be required for a good performance.

# State of the art of CW searches

Less candidates

Lower computing cost

Most searches are structured as follows:

**Main search stage** **Post-processing** **Application of vetoes** **Follow-up**

## Can we use machine learning to simplify this kerfuffle?

Typical searches run in a few million CPU hours on a supercomputer [Abbott+ PRD 100, 024004 (2019)].

GPU acceleration allows for deeper searches [Wette+ PRD 103, 083020 (2021)] and more diverse signals [Covas & Sintes PRD 99, 124019 (2019), Keitel & Ashton CQG. 35 (2018), 205003].

Broadest and deepest searches to date run on the volunteer-distributed computing platform Einstein@Home, consuming ~100 M CPU hours [Steltner+ ApJ 909 79 (2021)].

Extensive manual tuning, however, tends to be required for a good performance.

# Machine learning for the search of CWs

Most popular ML application: post-processing the results of a main search.

*Applied to real-data searches!

- *Identify signal tracks on a spectrogram using CNNs [Bayley+ PRD (2020)].
- Identify excess power using CNNs [Yamamoto+ PRD (2021), arXiv:2206.00882].
- Classification of candidate distributions using CNNs [Morawski+ Mach. Learn.: Sci. Technol. (2020)].
- Image-segmentation-based clustering using CNNs [Beheshtipour+ PRD (2020, 2021)].
- *Unsupervised clustering [Singh+ PRD (2017), Tenorio+ PRD (2021), Steltner+ arXiv:2207.14286].

To date, no "machine-learning method" has been used as a main stage in a CW search.

- Only proposal in this front: CNNs on Fourier transforms of the data [Dreissigacker+ PRD (2019, 2020)].

Can we design a machine learning pipeline that…

1. Does not require too much data pre-processing?
2. Is able to compete with current wide parameter-space searches?
3. Does not require 100 M CPU hours to run?

# Introducing Kaggle competitions

Kaggle: Online platform of data-analysis competitions.

- Host proposes a problem based on a dataset →participants propose ML-based solutions.

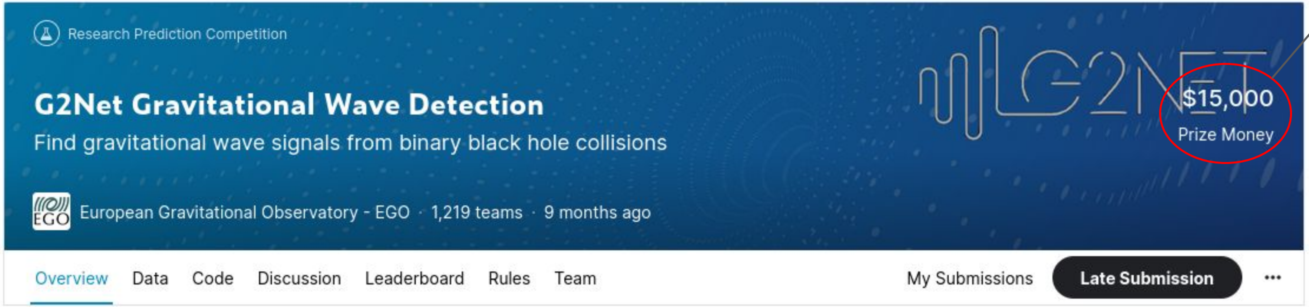
- Broad and diverse community of participants with different levels of expertise.

- Competitive nature:
  - Competitions tend to last for a few months.
  - Scores are made public in a global leaderboard.
  - Participants are encouraged to continuously improve their solution.

Upon completion of the challenge, winners are required to deliver their winning solutions in such a way that they can be fully reproduced.

# Previous GW data-analysis competitions

Monetary prize
for the winners!




Sponsored by Google.

People involved:
- Michael Williams
- Chris Messenger
- Elena Cuoco
- Chris Zerafa

Use ML to detect GW signals from binary black hole collisions.

- Over 1200 teams took part during the 3 months of competition.
- Diverse pool of expertise amongst participants, both at academic and industry levels.

Competition overview:

- Finite set of samples:
  - 500k training samples.
  - 220k testing samples.
- Solutions were ranking using the Area Under the ROC curve.
- No signal generation instructions were shared by the hosts.

# Towards a Kaggle competition to detect CWs

We are about to release a Kaggle competition aimed at detecting CWs.

- Profit from the momentum generated by the previous competition.
- CW are longer and weaker than CBC signals: More difficult challenge!
- CBC searches are close to optimal sensitivity; CW are limited by computing cost!
  - Plenty of room for improvement, considering the lack of ready-to-use techniques.

Lessons learnt from the previous challenge:

- Competitors benefit enormously from having as much data as possible.
  - Use PyFstat → New tutorials available!

- Knowledge is shared amongst competitors.
  - Extensive explanations are not required in order to have a successful competition.

# Summary

- Continuous gravitational waves are currently undetected forms of gravitational radiation.
- Typical searches for this kind of signals tend to be computationally limited.
- Little effort has been put into developing an end-to-end machine-learning search.
- We will use Kaggle as a mean to spark the development of ML-based CW searches, benefiting from the wide range of expertise amongst the Kaggle community.

# Acknowledgments

This work is supported by the Universitat de les Illes Balears; by European Union FEDER funds; the Spanish Ministry of Science and Innovation and the Spanish Agencia Estatal de Investigación grants PID2019-106416GB-I00/MCIN/AEI/10.13039/501100011033, RED2018-102661-T, RED2018-102573-E, FPA2017-90566-REDC; the Comunitat Autònoma de les Illes Balears through the Direcció General de Política Universitària i Recerca with funds from the Tourist Stay Tax Law ITS 2017-006 (PRD2018/24, PDR2020/11); the Conselleria de Fons Europeus, Universitat i Cultura del Govern de les Illes Balears; the Generalitat Valenciana (PROMETEO/2019/071); and EU COST Actions CA18108, CA17137.

Rodrigo Tenorio is supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (ref. FPU 18/00694).

UIB Universitat de les Illes Balears

GOIB CONSELLERIA FONS EUROPEUS, UNIVERSITAT I CULTURA DIRECCIÓ GENERAL POLÍTICA UNIVERSITÀRIA I RECERCA

una manera de fer europa

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Illes Sostenibles

IAC3 Institute of Applied Computing & Community Code.

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Unión Europea

Fondo Europeo
de Desarrollo Regional
"Una manera de hacer Europa"