Design of novel signal processing algorithms for the detection and reconstruction of gravitational waves emitted by core-collapse supernovae

A Data Management Plan created using DMPonline.be

Creator: Milan Wils

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 11POK24N

ID: 206759

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Since the first detection of gravitational waves (GWs) by the LIGO detectors of a binary black hole merger in 2015, many more binary mergers have been detected. A new generation of GW detectors will be build, whose improved sensitivity will allow the detection of core-collapse supernovae (CCSNe). Those detections can result in a better understanding of CCSNe, interaction between the fundamental forces of nature in extreme environments and astrophysical processes that depend on CCSNe such as nucleosynthesis. New algorithms are required because CCSNe waveforms are estimated through various simulations and are stochastic. The proposal aims to develop these new algorithms by applying state-of-the art digital signal processing principles. The null-stream analysis will be repurposed to detect GWs with low latency and minimal assumptions on the source while deterministic section of the GW will be targetted by a template-based search. The maximal amount of information will be extracted from the waveform by exploiting the multi-messenger structure of CCSNe. A generalised sidelobe cancellation (GSC) structure will attempt to exploit the information in the null-stream for noise reduction. An optimal wavelet transformation will be selected to improve the reconstruction of the GW. Extensive simulations will be performed to investigate the maximal source distance of the proposed algorithms and compare them to the state-of-the-art.

Last modified: 29-04-2024

Design of novel signal processing algorithms for the detection and reconstruction of gravitational waves emitted by core-collapse supernovae

FWO DMP (Flemish Standard DMP)

---

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
LVK O4	Gravitational wave strain data from the fourth observing run, including calibration uncertainty	Reuse existing data	Digital	Observational	.gwf, .h5, .txt	< 50 TB	
ET MDC	Mock Data Challenge for the Einstein Telescope	Generate new data	Digital	Simulation data	.gwf	< 1 TB	
cWB	coherent WaveBurst	Reuse existing data	Digital	Software	.C, .h	< 100 MB	
cWB Analyses	coherent WaveBurst analysis setups and results	Generate new data	Digital	Compiled/aggregated data	.root, .h5, .html, .png, .txt	< 1 TB	
Python Scripts	Packages that are developed, outreach demos, data visualisation scripts etc.	Generate new data	Digital	Software	.py	< 100 MB	
Tensorlab	Matlab/Python toolbox for tensor methods	Reuse data	Digital	Software	.m, .py	< 100 MB	
Supernova waveforms	Simulated 2D and 3D gravitational wave signatures of core-collapse supernovae	Reuse data	Digital	Simulation data	.csv, .txt, .dat, .h5	< 100 MB	
ET Supernova search pipeline	Unmodelled search pipeline for the Einstein Telescope	Generate new data	Digital	Software	.cpp, .hpp, .py	< 100 MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

- LVK O4 Data
  - Proprietary data of the LIGO-Virgo-KAGRA collaboration
  - Relevant data will be made public on the Gravitational Wave Open Science Center (GWOSC) on 23/05/2026
- cWB
  - Latest version is proprietary to the LIGO-Virgo-KAGRA collaboration
  - Current public version (O3)
- Tensorlab
  - Latest version
- Supernova waveforms
  - Not a uniform dataset but a compilation of 10-20 published waveforms in the literature. The compilation of this list is part of the research project.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

Any analyses results that are the result of a non-public version of cWB and/or non-public LVK data have to be approved by the LVK Collaboration prior to publication.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

The generated datasets will be accompanied by a readme file that specifies assumptions made during its generation, as well as a link to the code repository that was used to generate it. Standard data formats such as Gravitational-Wave Frames (.gwf), HDF5 and ROOT will be used to avoid ambiguity in the data specification.
All generated code will be documented through descriptions in comments. For large codebases such as the unmodeled search pipeline, doxygen will be used to automatically generate documentation. All, packages will include installation instructions through a readme file.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The simulated data will store the metadata in the standardised header of .gwf files.

3. Data storage & back-up during the research project

Where will the data be stored?

All generated code will be stored on Gitlab repositories. Depending on the use case and collaborators, this can any of the following servers:

- gitlab.et-gw.eu
- git.ligo.org
- gitlab.kuleuven.be
- gitlab.esat.kuleuven.be

The generated dataset will be stored temporarily on collaboration computing clusters. After publication, the software to re-generate the data will be made public. The software version and used settings will be provided as well to ensure reproduction capability.

How will the data be backed up?

All Gitlab repositories are follow an internal backup procedure (KU Leuven servers, IGWN Computing)

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

- The size of the code repositories is negligible compared to common requirements on Gitlab servers.
- The ET-MDC dataset and cWB analysis results are well within the accepted limit of the LVK CIT cluster limits

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Codes stored on the KU Leuven servers are protected through MFA.
All collaboration resources are protected through authorisation with the collaboration account. The LVK collaboration is also transitioning to MFA.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

Because all code / generated data will be hosted on collaboration resources or KU Leuven Gitlab servers, there is no direct cost associated with its storage.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data will be preserved for 10 years, according to the KU Leuven RDM policy. The exception is the Einstein Telescope Mock Data where only the software and its configuration will be stored due to its large volume (1 TB) and reproducibility (<1 day on a laptop).

Where will these data be archived (stored and curated for the long-term)?

All generated code will be stored on Gitlab repositories. Depending on the use case and collaborators, this can any of the following servers:

gitlab.et-gw.eu
git.ligo.org
gitlab.kuleuven.be
gitlab.esat.kuleuven.be

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

There are no direct costs because all data will be stored on collaboration resources or a KU Leuven Gitlab server.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- No (closed access)
- Yes, in an Open Access repository

- All code developped during this project will be made public as soon as it is stable and the first results have been published.
- cWB simulation results will not be made public because they are the result of proprietary data.
- The configuration of the ET MDC will be made public.

If access is restricted, please specify who will be able to access the data and under what conditions.

The cWB analysis results will be accesible to all LVK members

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Intellectual Property Rights

All products derived from non-public versions of LVK Collaboration code and / or data are subject to approval before they can be made public.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

All code will be made available through their respective git repositories. The ET MDC data configuration will be available in the first published paper.

When will the data be made available?

Upon publication of the results

Which data usage licenses are you going to provide? If none, please explain why.

Creative Commons Attribution (CC-BY)

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

The initial publications will serve as the citable

What are the expected costs for data sharing? How will these costs be covered?

There are no direct costs because all data will be stored on collaboration resources or a KU Leuven Gitlab server.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Milan Wils

Who will manage data storage and backup during the research project?

Milan Wils

Who will manage data preservation and sharing?

Milan Wils

Who will update and implement this DMP?

Milan Wils