Assurer la qualité logicielle pour une recherche réproductible

Philip DAVIS
on behalf of the LuTH-Caen group in Virgo :
LPC-Caen, GANIL, LuTH, Observatoire de Strasbourg

In collaboration with : University of California, Fullerton

Journées Informatiques 2024

cnrs ENSI CAEN UNICAEN UNIVERSITÉ CAEN NORMANDIE Normandie Université

LPC Caen

Background Context

Neutron stars (NSs) : dense and compact objects formed from progentors with masses ≥ 8-10 Msun :

● Mass : ~ 1-2 Solar masses

● Radii : ~10 – 14 km

● Mean density : ~10^{14} – 10^{15} g/cm^3

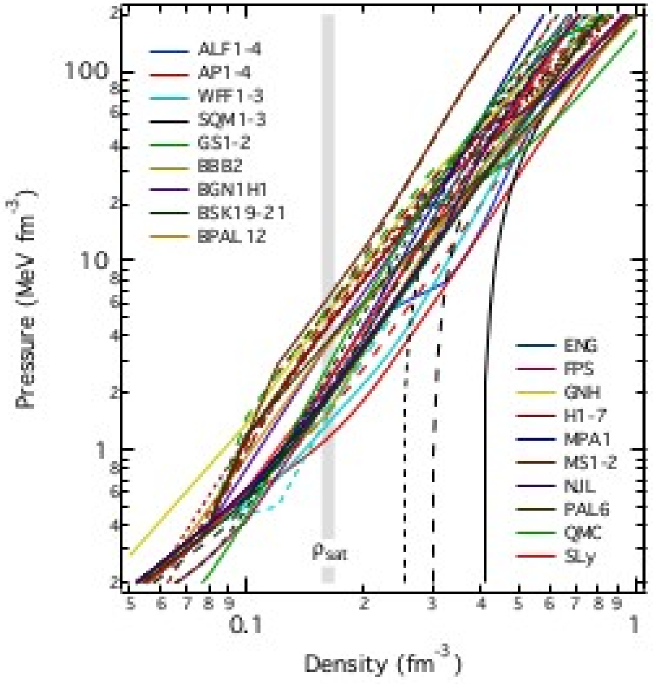
For cold, mature NSs :

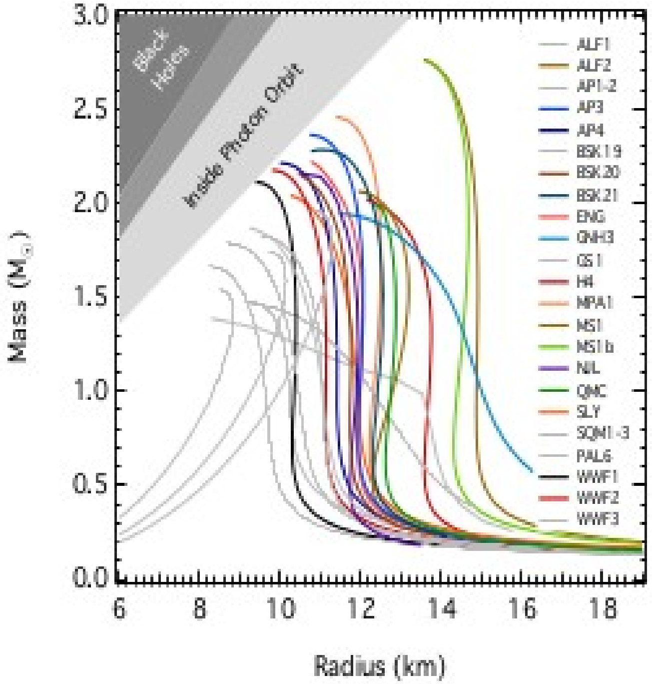
● The equation of state (EoS) describes a relationship between the pressure and the density,

● Once we have the EoS, we can determine the structure of a NS

● Then calculate global NS properties, e.g. mass, radius, tidal deformability.

Credit : Ozel & Freire 2016





There are many NS EoS,

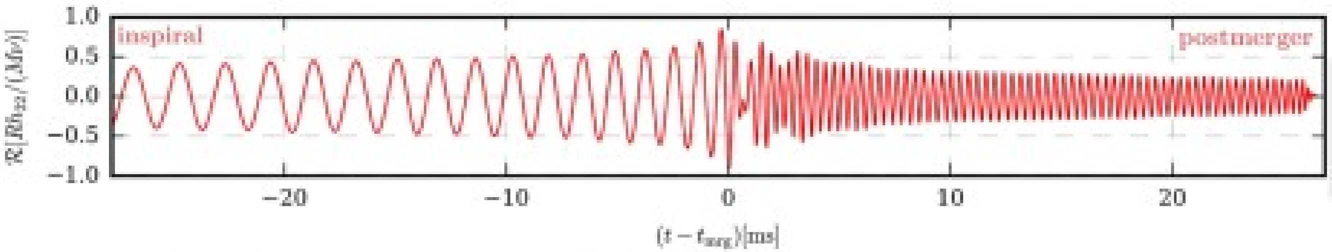
→ Comparison between calculated and observed NS properties can help constrain the EoS.

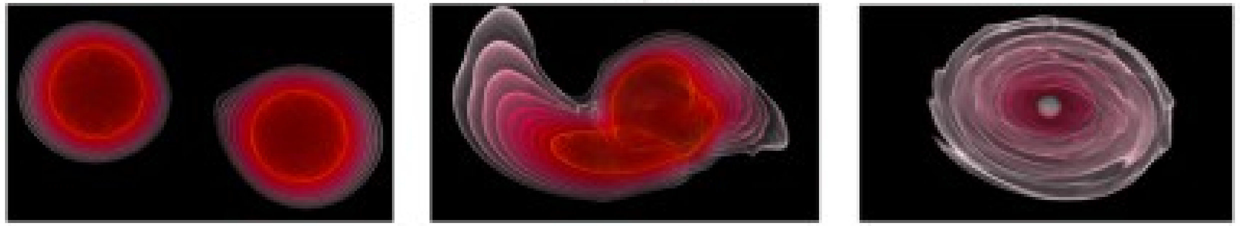
Background Context (cont.)

Credit : Virgo/ Ligo collaboration

Terrestrial experiments can only probe low-density regimes. For higher densities, require astrophysical constraints, e.g. gravitational waves (GWs) during NS merger events.

Credit : Dietrich et al. 2021





Pre-merger Merger Post-merger

Development of numerical tools for the LIGO-Virgo-Kagra collaboration by members of LuTH-Caen group in Virgo (LPC, LuTH, Strasbourg, GANIL).

The shape of the GW signal (the « waveform ») depends on the property of matter and hence the EoS of the NS.

Comparing observed and simulated waveforms, provides information on the EoS.

Aim of role : raise awareness of good software development practices and data management so that quality numerical tools and data can be used by the scientific community.

Why is this important ?

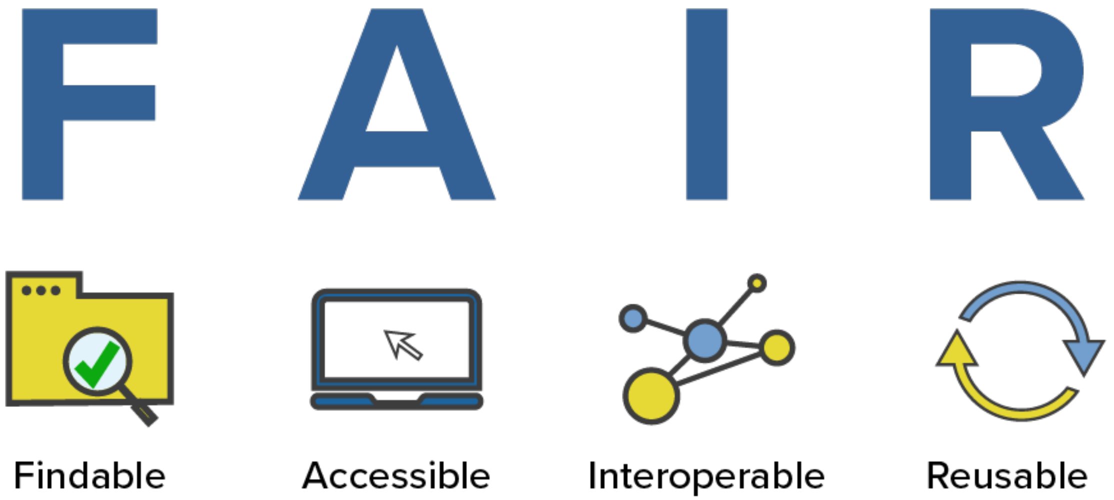
Good software and data practices ensure :

- Robustness and reliabability,
- Transparency,
- Traceability,
- Reproducibility.

These are in turn important for the Open Science initiative, which :

- Improves visibility of research,
- Ensures research can be validated,
- Improves collaborations,
- Reduces duplicated effort.

Credit : www.nlm.nih.gov



Challenges for reproducible science

Obstacles to reproducible research :

- Code/ Data cannot be found,
- Code/ Data is not publicly accessible,
- Software contains bugs and poorly documented,
- Missing or incompatible libraries/ dependencies.

https://www.nature.com/articles/533452a

1,500 scientists lift the lid on reproducibility

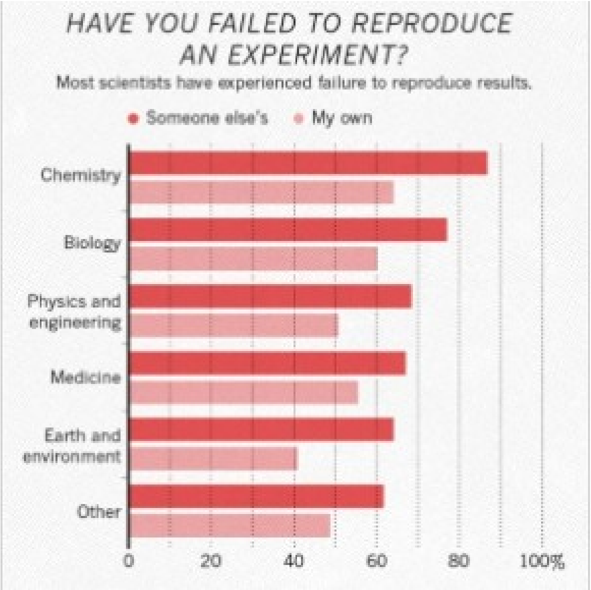
Monya Baker

Nature **533**, 452–454 (2016) | Cite this article

This article has been updated

Survey sheds light on the 'crisis' rocking research.

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.



Raising awareness of good practices

Created training material covering software development working practices :

- Version control with git and GitLab,
- Design and documentation,
- Tests,
- Code reviews.

First course given to LuTH-Caen group members Autumn 2022. Recieved positive feedback.

Second course, April 2024 (to be given yearly)

Next steps : to include material on publishing code and data (choosing a license, writing documentation, software environment etc)

Software Development and Testing Tutorial

2. Development working practices

Philip DAVIS
Ingenieur de Recherche
Laboratoire de Physique Corpusculaire

Recap

Code repository: A database, usually located on an external server, where code is stored. GitLab is just one such example.

Version control: A way to track code changes in a organized and logical way.

Code branch: A separate copy of the latest, stable version of the code where we can make changes without interfering with the Master. Changes to the code are saved via **commits.**

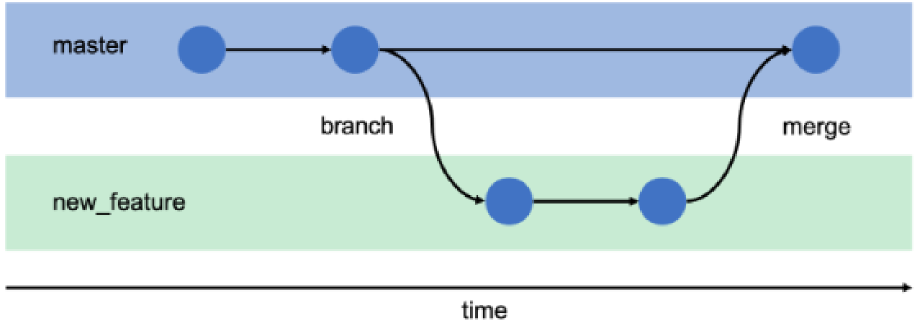


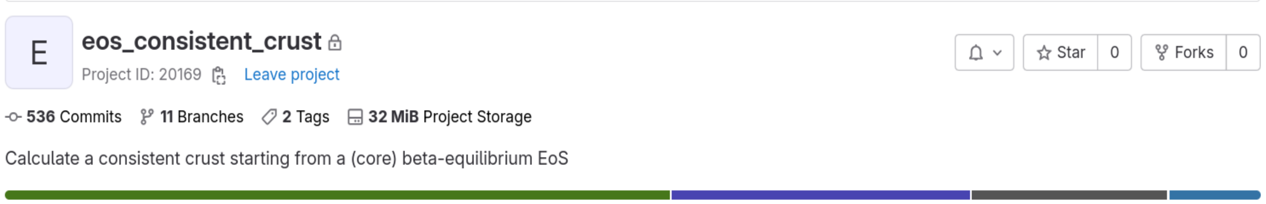
Fig. 1: Illustration of a code branch (credit: gitbookdown.dallasdatascience.com)

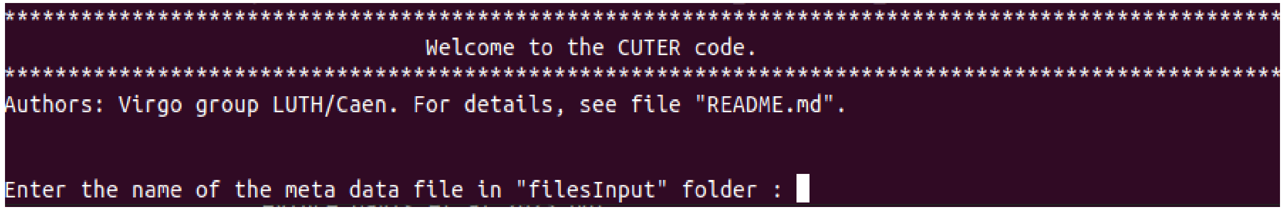
Material located at :
https://gitlab.in2p3.fr/eos-for-virgo/training

CUTER (Crust Unified Tool for Equation-of-State Reconstruction)

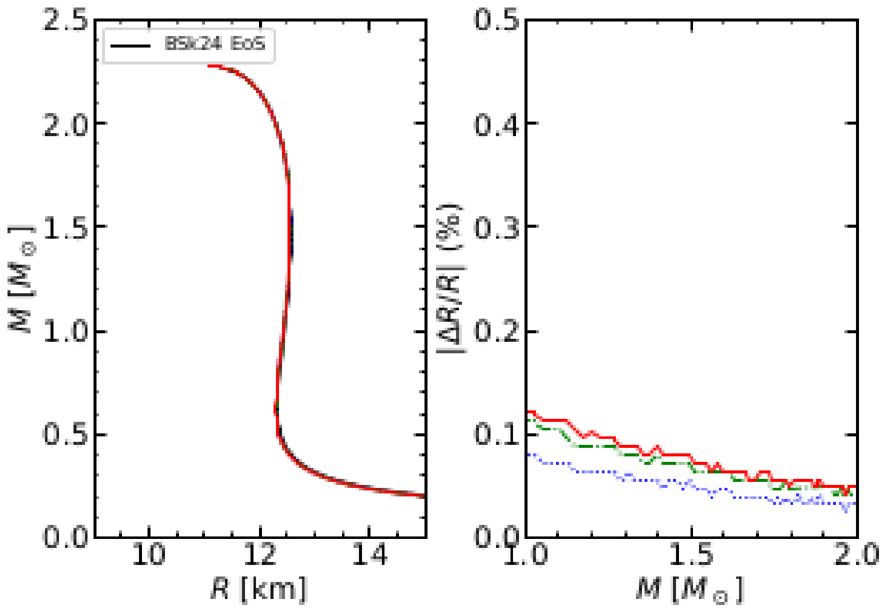
- A NS equation of state is needed to relate different global NS properties, for example mass and radius,

- Inconsistent treatment of NS crust (« non-unified ») can introduce biases.





Davis et al. 2024



Numerical tool, CUTER, developed for the LIGO-Virgo-Kagra collaboration :

- Attach a thermodynamically consistent crust to an equation of state describing the core of a NS,

- Aim : reduce errors of inferred global properties of NS. Important for the next generation of gravitational wave detectors (e.g. Einstein Telescope).

CUTER (cont.)

Working practices followed :

- Software development « workflow »
(issue -> branch -> testing -> review -> merge -> close issue)

- Hosted on IN2P3 and LIGO GitLab sites,

- Documentation (e.g. README),

- Addition of a License (protect rights of both the users and developers),

- Sign-off by external reviewer,

- `Conda` to manage software environment,

- `Pytest` for automated testing.

`CUTER` was opened to the LIGO-Virgo-Kagra collaboration May 2023. Paper describing the code published in A&A (Davis et al. 2024).
https://doi.org/10.1051/0004-6361/202348402

`CUTER` also published on Zenodo.
DOI : 10.5281/zenodo.10781539

Since then :

- `CUTER V2` released July 2024 (outer crust reconstruction),

- `CUTER V3` under development (Bayes tools) (forseen spring-summer 2025)

Future plans

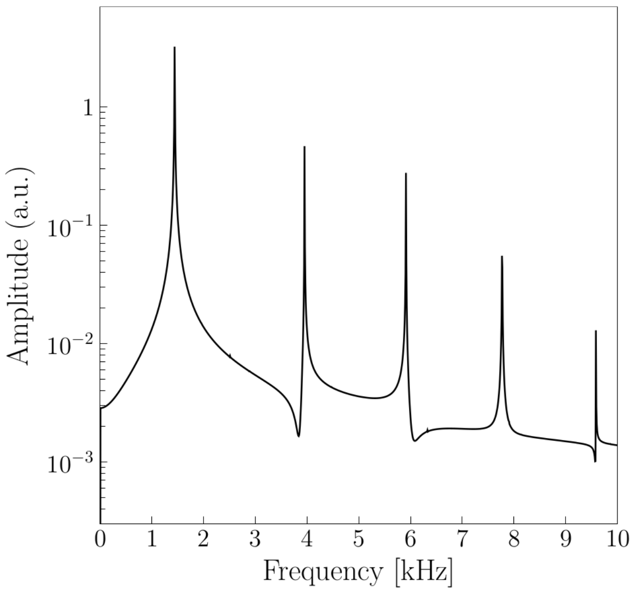
Medium term (3-4 years)

- Parametric study of NSs with a large set of microphysical inputs, simulating possible GW signals, from a post-merger of NS+NS binary,
- In collaboration with researchers from the Observatoire de Strasbourg, in the framework of the *ANR GW-HNS (2023-2025) project,*
- Development of hydro code to model oscillating NSs and analytical representations of EoS underway.

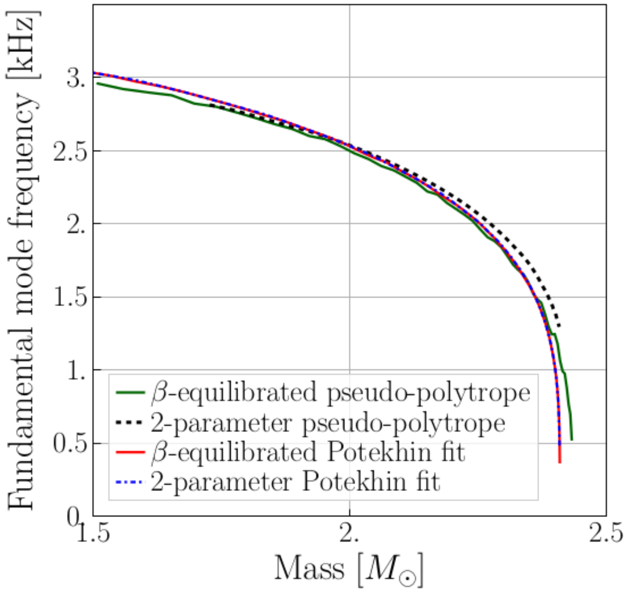
Long term (> 5 years)

- Provide data (e.g. GW waveforms) for the LIGO-Virgo-Kagra analysis pipelines and other collaborators,

Servignat et al. 2023, Class. Quantum Grav. DOI : 10.1088/1361-6382/acc828



Servignat (+Davis) et al. 2023, Phys. Rev. D (DOI: https://doi.org/10.1103/PhysRevD.109.103022

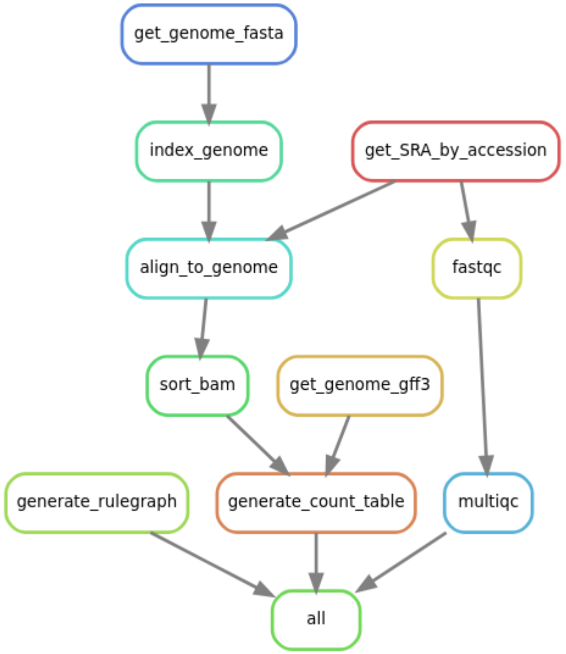


Associated challenges

- Storing, sharing and publication of large datasets,
- Transparency regarding the provenance of data sets (e.g. processing performed, code versions) → meta data ?
- Computationally intensive simulations,
- Managing software environments for multiple langages (`Python, C/C++, Fortran`)
- Managing and automating complex task « workflows ».

Make use of publicly available tools, e.g.

- `Datalad`, `Git/Gitlab` : management of code and data,
- `Docker, Conda` : management of software environments,
- `Snakemake` : automating task workflows.



Credit : nbis-reproducible-research.readthedocs.io

Merci pour votre attention