Assuring Quality Software for Reproducible Research

MAC meeting, Orsay

30th May 2024

Philip Davis

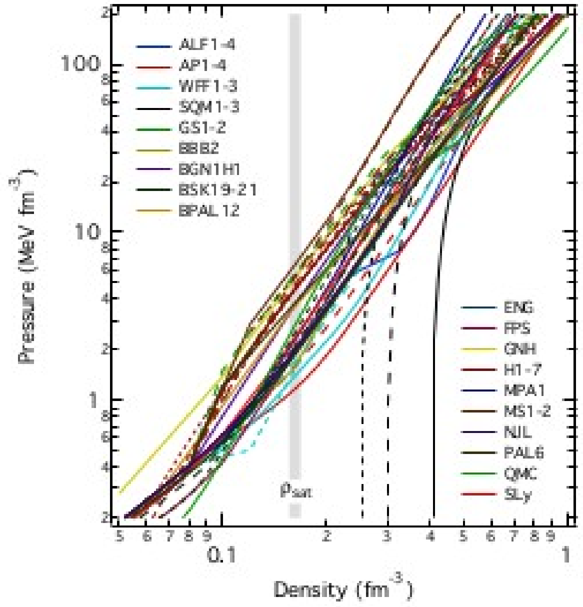
LPC Service Informatique/ Theory Group

In collaboration with Observatoire Astronomique de Strasbourg, California State University – Fullerton, GANIL

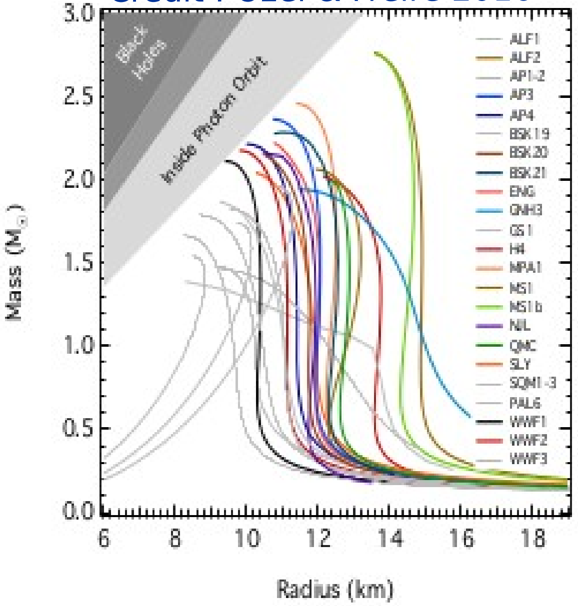
Background Context

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



Credit : Ozel & Freire 2016



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.

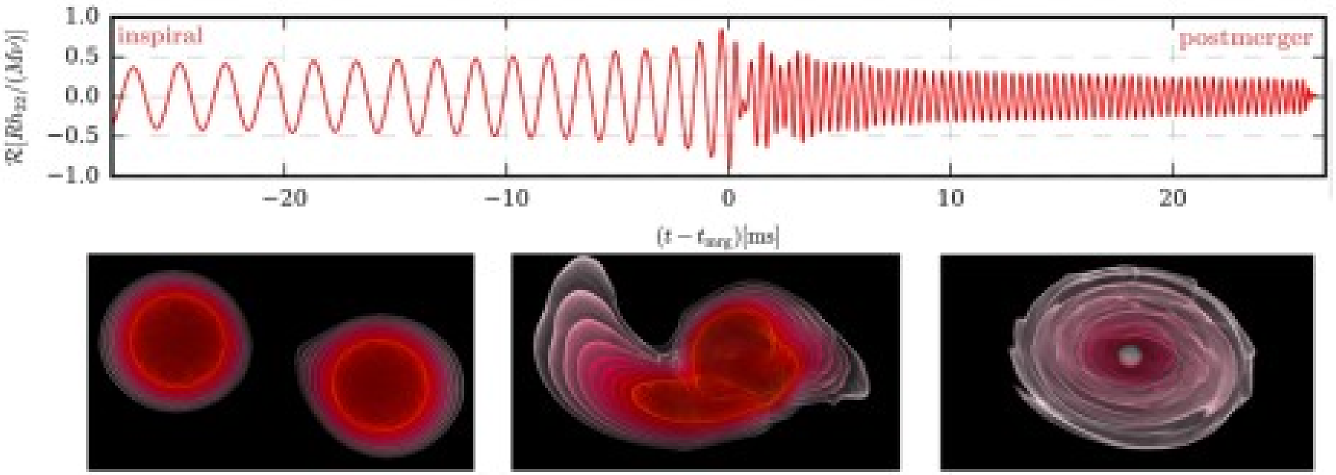
There are many NS EoS,

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

Background Context (cont.)

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

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.

Credit : Virgo/ Ligo collaboration

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

Background Context (cont.)

Professional context :

- PhD astrophysics (2005-2009),
- Post-doctoral position, Brussels (2010-2015),
- Scientific Software Engineer (2015-2022).

→ Background in software development, scientific research and high-performance computing

Currently

- Ingénieur de Recherche (CDI) since April 2022,
- Member of the « LuTH-Caen » group,
- 1 ETP Virgo collaboration.

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.

FAIR :

Findable
Accessible
Interoperable
Reproducible

Credit : Normadie Université

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

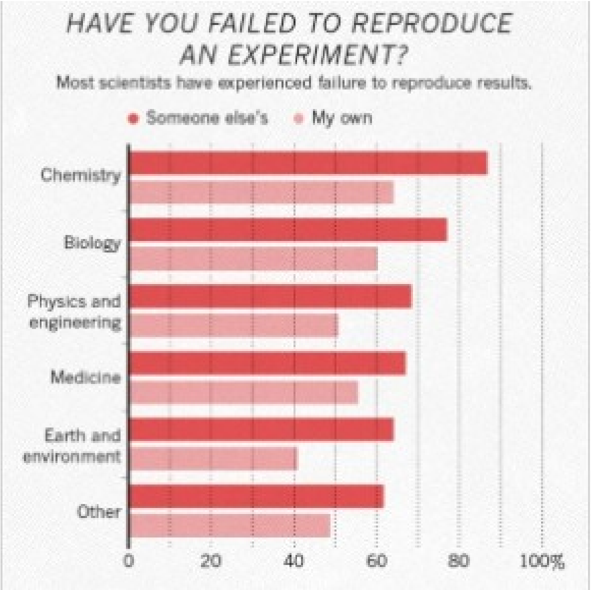
Monya Baker

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

Software Development and Testing Tutorial

2. Development working practices

Philip DAVIS
Ingenieur de Recherche
Laboratoire de Physique Corpusculaire

IN2P3

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.**

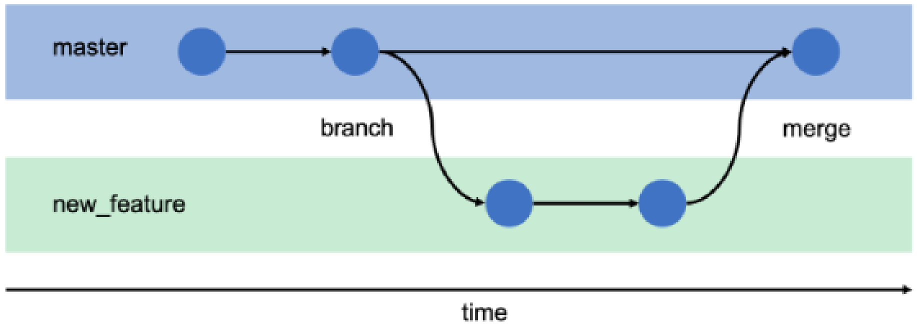


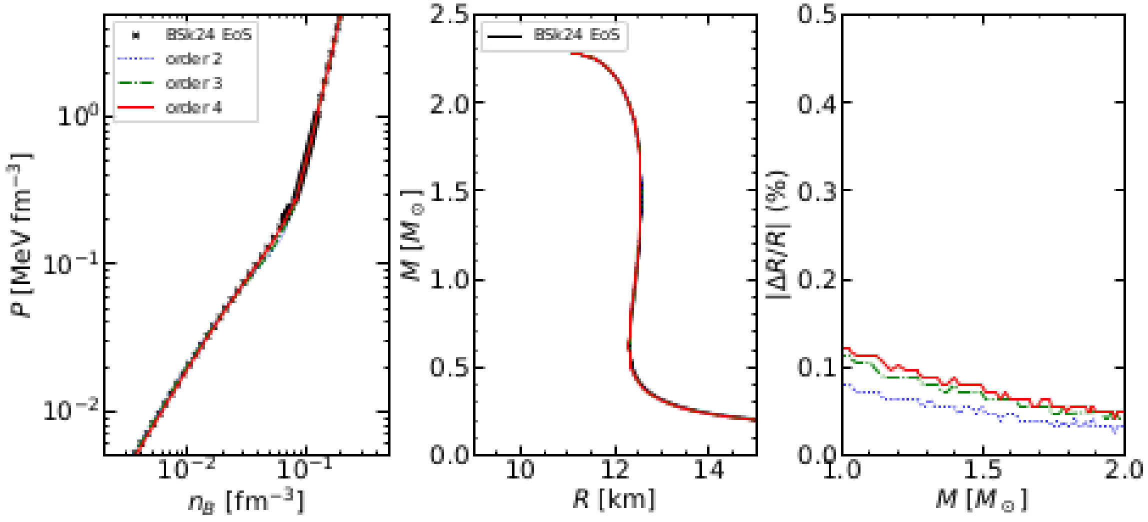
Fig. 1: Illustration of a code branch (credit: gitbookdown.dallasdatascience.com)

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

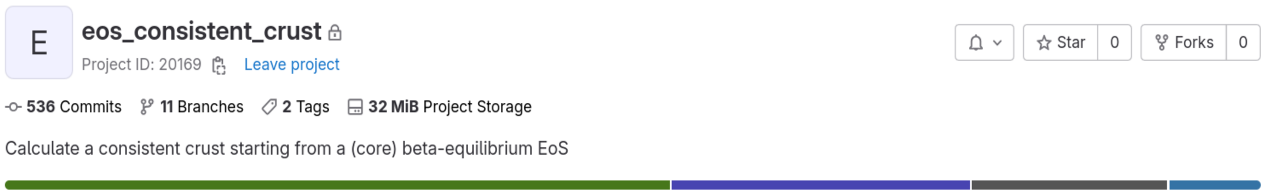
CUTER (Crust Unified Tool for Equation-of-State Reconstruction)

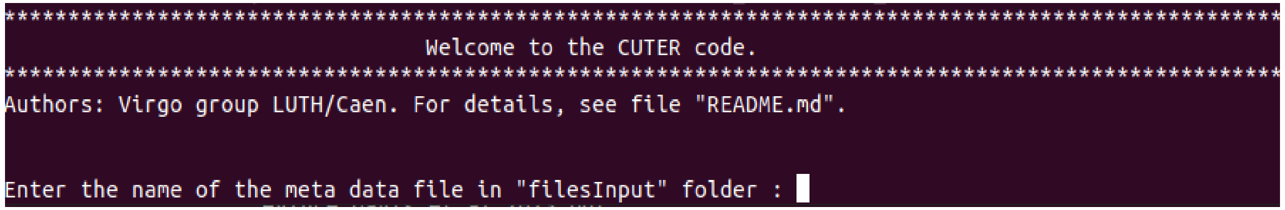
- A NS equation of state is needed to relate different global NS properties, for example mass and radius,

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



Credit : Davis et al. 2023 (accepted, A&A, in press)





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 »,
- Hosted on IN2P3 and LIGO GitLab sites,
- Documentation (e.g. README),
- Addition of a License (protect rights of both the users and developers),
- Internal reviews + sign-off by external reviewer,
- `Conda` to manage software environment,
- `Pytest` for automated testing.

CUTER was opened to the LIGO-Virgo collaboration for the O4 observational campaign. Paper describing the code has been accepted by A&A (Davis et al. 2023, in press).
https://doi.org/10.1051/0004-6361/202348402

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

Next steps :

- Additional developments for O4b campaign currently under review : addition of a consistent outer crust only,
- Tools to perform a Bayesian analysis,
- Include CUTER in `LaLSuite` (analysis pipeline).

Future plans

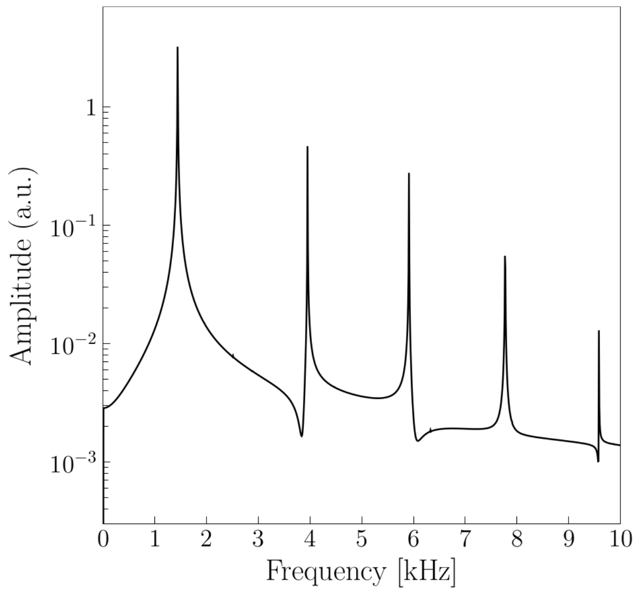
Medium term (3-4 years)

- 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.

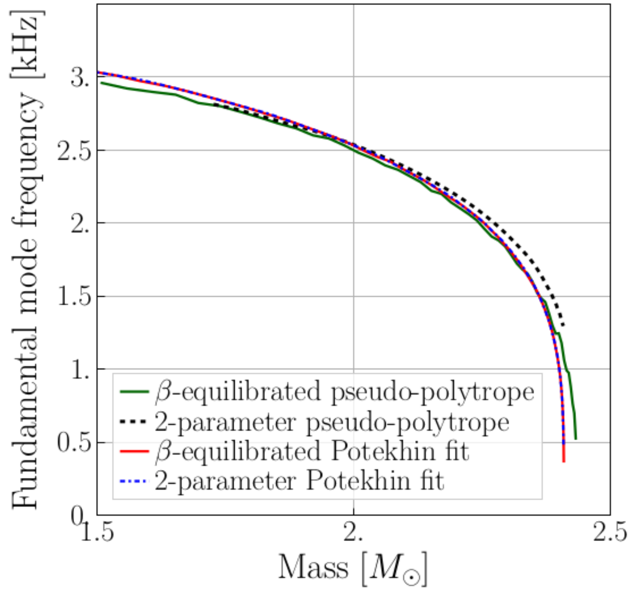
Long term (> 5 years)

- Provide data (e.g. GW waveforms) for the Ligo-Virgo 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

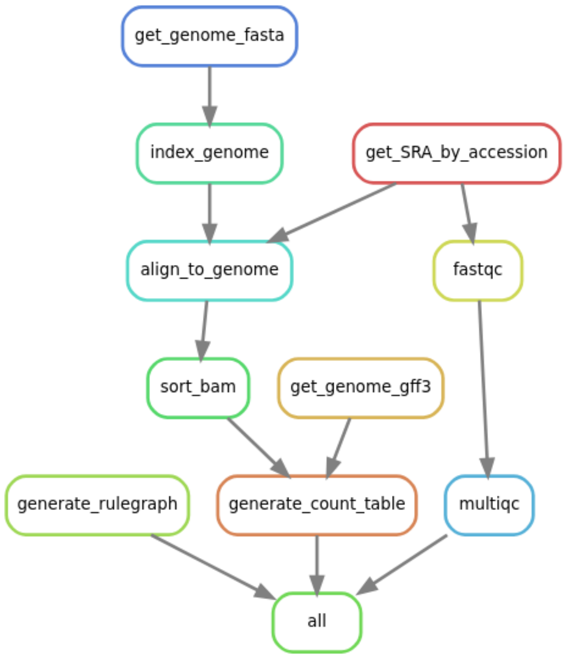


Associated challenges

- 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

EoS reconstruction (1)

- Start from a (high-density) beta-equilibrium EoS $\mathcal{E}_\beta(n_B)$
  --> corresponding energy per baryon of homogeneous nucleonic matter

$$e_{\mathrm{nuc},\beta}(n_B) = \frac{1}{n_B}\left[\mathcal{E}_\beta(n_B) - \mathcal{E}_e(n_e(n_B)) - n_e m_p c^2 - (n_B - n_e) m_n c^2\right]$$

see also Essick et al. PRC 104, 065804 (2021) – up to 2nd order ($K_{\mathrm{sat,sym}}$)

$$e_{\mathrm{nuc}}(n_B, \delta) = e_{\mathrm{is}}(n_B) + e_{\mathrm{iv}}(n_B)\delta^2 \qquad \delta = \frac{n_n - n_p}{n_B}$$

$$e_{\mathrm{is}}(n_B) = E_{\mathrm{sat}} + \frac{1}{2}K_{\mathrm{sat}}x^2 + \frac{1}{3!}Q_{\mathrm{sat}}x^3 + \ldots \qquad x = \frac{n - n_{\mathrm{sat}}}{3n_{\mathrm{sat}}}$$

$$e_{\mathrm{iv}}(n_B) = E_{\mathrm{sym}} + L_{\mathrm{sym}}x + \frac{1}{2}K_{\mathrm{sym}}x^2 + \frac{1}{3!}Q_{\mathrm{sym}}x^3 + \ldots$$

from $\mathcal{E}_\beta(n_B)$ + β equilibrium condition + knowledge of isoscalar part
--> one can extract isovector part --> **“inversion” procedure**

EoS reconstruction (2)

- Start from a (high-density) beta-equilibrium EoS $\mathcal{E}_\beta(n_B)$

  ⇒ corresponding energy per baryon of homogeneous nucleonic matter

$$e_{\mathrm{nuc},\beta}(n_B) = \frac{1}{n_B}\left[\mathcal{E}_\beta(n_B) - \mathcal{E}_e(n_e(n_B)) - n_e m_p c^2 - (n_B - n_e) m_n c^2\right]$$

- Meta-model approach for the energy density of nucleonic matter

  (Margueron et al., PRC 97, 025805 (2018))

$$\delta = \frac{n_n - n_p}{n_B}$$

$$x = \frac{n - n_{\mathrm{sat}}}{3 n_{\mathrm{sat}}}$$

$$e_{\mathrm{nuc}}(n_B, \delta) = t^\star_{\mathrm{FG}}(n_B, \delta) + e_{\mathrm{is}}(n_B) + e_{\mathrm{iv}}(n_B)\delta^2$$

$$e_{\mathrm{is,iv}}(n_B) = \sum_{k=0}^{N} \frac{v_k^{\mathrm{is,iv}}}{k!} x^k u_k(x)$$

❖ kinetic term with *m**

❖ *N* > 2 ⇒ beyond parabolic approx.

⇒ important for correct calculation of the crust properties

(Dinh Thi et al., A&A 654, A114 (2021))

zero-density limit

E_{sat}, K_{sat}, Q_{sat}, …; E_{sym}, L_{sym}, K_{sym}, Q_{sym}, …

"inversion procedure" as in Mondal & Gulminelli, PRD 105, 083016 (2021)

Crust (Unified) Tool for Equation-of-state Reconstruction : CUTER

- $\mathcal{E}_\beta(n_B)$ Inputs: - (at high-density from ~ $n_{\rm sat}$)
  - isoscalar parameters $n_{\rm sat}$, $E_{\rm sat}$, $K_{\rm sat}$ ($Q_{\rm sat}$, $Z_{\rm sat}$)

- Perform "inversion" procedure --> $E_{\rm sym}$, $L_{\rm sym}$, $K_{\rm sym}$ ($Q_{\rm sym}$, $Z_{\rm sym}$)

- Calculate the crust EoS using a compressible-liquid drop model (CLDM) for ions
  (Carreau et al., EPJA 55, 188 (2019); Dinh Thi et al., A&A 654, A114 (2021))

$$\mathcal{E}_{\rm inhomog} = \mathcal{E}_e + \mathcal{E}_{\rm g}\left(1 - \frac{V_N}{V_{\rm WS}}\right) + \frac{E_i}{V_{\rm WS}}$$

$$E_i = M_i c^2 + E_{\rm bulk} + E_{\rm Coul} + E_{\rm suf+curv}$$

- Match low-density consistent part with original high-density EoS

N.B.: (i) inversion done chosing N+1 points around saturation
--> only nucleonic degrees of freedom
(ii) for the bulk part of ions and gas --> same meta-model --> unified
(iii) if isoscalar parameters not known --> assigned (using BSk24 param.)
– or they can be picked at random

Code available to the LIGO-Virgo-KAGRA collaboration (GitLab)