Asimov: A framework for coordinating parameter estimation workflows

Daniel Williams[1], John Veitch[1], Maria Luisa Chiofalo[4], Patricia Schmidt[6], Richard P. Udall[5], Avi Vajpeji[2, 3], and Charlie Hoy[7]

1 School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, United Kingdom **2** School of Physics and Astronomy, Monash University, Clayton VIC 3800, Australia **3** OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton VIC 3800, Australia **4** Department of Physics "Enrico Fermi," University of Pisa, and INFN, Largo Bruno Pontecorvo 3 I-56126 Pisa, Italy **5** LIGO Laboratory, California Institute of Technology **6** School of Physics and Astronomy and Institute for Gravitational Wave Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 9TT, United Kingdom **7** Cardiff University, Cardiff CF24 3AA, UK

Summary

Since the first detection in 2015 of gravitational waves from compact binary coalescence (B. P. Abbott & others, 2016a), improvements to the Advanced LIGO and Advanced Virgo detectors have expanded our view into the universe for these signals. Searches of the of the latest observing run (O3) have increased the number of detected signals to 90, at a rate of approximately 1 per week (The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, et al., 2021; The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al., 2021). Future observing runs are expected to increase this even further(B. P. Abbott & others, 2020). Bayesian analysis of the signals can reveal the properties of the coalescing black holes and neutron stars by comparing predicted waveforms to the observed data (B. P. Abbott & others, 2016b). The proliferating number of detected signals, the increasing number of methods that have been deployed (Ashton & others, 2019; Lange et al., 2018; Veitch & others, 2015), and the variety of waveform models (Khan et al., 2020; Ossokine & others, 2020; Pratten & others, 2021) create an ever-expanding number of analyses that can be considered.

`Asimov` is a python package which is designed to simplify and standardise the process of configuring these analyses for a large number of events. It has already been used in developing analyses in three major gravitational wave catalog publications (R. Abbott & others, 2021; The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, et al., 2021; The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al., 2021). The source code of `Asimov` is archived to Zenodo (Williams et al., 2021).

Statement of Need

While these developments are positive, they also bring considerable challenges. The first of these lies with the high rate at which gravitational waves can now be detected; thanks to the improved sensitivity of the detectors they observe a much larger volume of space, and the increasing size of the detector network has also increased the total time during which observations occur. The second comes from developments in the analysis techniques and related software. Development of these techniques has accelerated in a short period of time, and the landscape of analysis software has become diverse. It is desirable to be able to use

41 these techniques with ease, but thanks to the highly distributed development process which
42 has produced them, they often have highly heterogeneous interfaces.

43 We developed asimov as a solution to both of these problems, as it is capable both of organising
44 and tracking a large number of on-going analyses, but also of performing setup and post-
45 processing of several different analysis pipelines, providing a single uniform interface. The
46 software has been designed to be easily extensible, making integration with new pipelines
47 straight-forward.

48 In addition, ensuring that the large number of analyses are completed successfully, and their re-
49 sults collated efficiently proved a formidable challenge when relying on "by-hand" approaches.
50 The LIGO Scientific Collaboration operate a number of high-throughput computing facilities
51 (the LIGO Data Grid [LDG]) which are themselves controlled by the `htcondor` scheduling
52 system. `asimov` monitors the progress of jobs within the `htcondor` ecosystem, resubmits
53 jobs to the cluster which fail due to transient problems, such as file I/O errors in computing
54 nodes, and detects the completion of analysis jobs. Upon completion of a job the results are
55 post-processed using the PESummary python package (Hoy & Raymond, 2021), and humans
56 can be alerted by a message posted by `asimov` to a Mattermost or Slack channel. Interaction
57 with `htcondor` will also allow jobs to be submitted to the Open Science Grid in the future.

58 Prior to the development of `asimov` analyses of gravitational wave data had been configured
59 and run manually, or had relied on collections of shell scripts. `Asimov` therefore constitutes
60 an new approach, designed to be both more maintainable, and to improve the reproduciblity
61 of results generated by analysis pipelines.

Implementation

63 In order to produce a uniform interface to all of its supported pipelines, `asimov` implements a
64 YAML-formatted configuration file, which is referred to as its "production ledger." This file is
65 used to specify the details of each event to be analysed, details about the data sources, and
66 details of each pipeline which should be applied to the specified data. This allows identical
67 settings to be used with multiple different pipelines, with a minimum of configuration, reducing
68 the possibility of transcription errors between setups. In the current implementation of `Asimov`
69 the production ledger is stored using an issue tracker on a custom `Gitlab` instance, with each
70 issue representing a different event. This approach is, however, neither flexible nor scalable,
71 and future development will use an alternative means of storing the ledger.

72 `Asimov` simplifies the process of gathering and collating the various settings and data-products
73 required to configure an analysis. These include data quality information: data from gravi-
74 tational wave detectors can be affected by non-stationary noise or "glitches" which must be
75 either be removed before analysis, or the analysis must be configured to mitigate their effect
76 on final results. These data are provided to `Asimov` in YAML format from the appropriate
77 team, and used to make appropriate selections in the analysis.

78 The analysis of gravitational wave data is generally performed within a Bayesian framework,
79 which requires prior probability distributions being chosen before the analysis. Ideally these
80 distributions would be chosen such that a very broad range of parameter values is explored and
81 sampled, however this is computationally impractical, and to improve the speed and efficiency
82 of the analysis a rough guess of the parameters is required. This is normally determined by
83 "preliminary" analyses, rougher, rapid analyses performed, which are themselves informed by
84 the detection process which identified the event in the raw detector data. These prior data
85 are analysed by the PEConfigurator tool to determine appropriate prior ranges, and settings
86 for the waveform approximant to be used in the analysis.

87 The calibration of the detectors; the correspondance between the strain on the detector and the
88 intensity of light at the interferometer's exit port, can change over the course of an observing

run. The uncertainty in this quantity is marginalised by many of the analyses, which requires data files to be collected and provided to the analyses.

Once the correct data, settings, and calibration information has been identified and collected it is possible to configure analyses. `Asimov` allows analyses to be described as a dependency tree, allowing the output data products from one analysis to be used as an input for another. This is often useful for coordinating the determination of the PSD of the analysed data.

Each pipeline is configured with a mixture of configuration files and command-line arguments. `Asimov` produces the appropriately-formatted configuration file for each pipeline using a template and substitutions from the production ledger. The appropriate command line program is then run for the given pipeline, in order to produce an execution environment and submission data for the `htcondor` scheduling system. This is then submitted to the LDG, and the job id is collected and stored by `asimov`.

It is then possible to automatically monitor the progress of jobs on the LDG, produce a webpage summarising the status of all on-going analyses, and detect the completion of jobs and initialise post-processing.

Acknowledgements

References

Abbott, B. P., & others. (2016a). Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, *116*(6), 061102. https://doi.org/10.1103/PhysRevLett.116.061102

Abbott, B. P., & others. (2016b). Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, *116*(24), 241102. https://doi.org/10.1103/PhysRevLett.116.241102

Abbott, B. P., & others. (2020). Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, *23*(1), 3. https://doi.org/10.1007/s41114-020-00026-9

Abbott, R., & others. (2021). GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X*, *11*, 021053. https://doi.org/10.1103/PhysRevX.11.021053

Ashton, G., & others. (2019). BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, *241*(2), 27. https://doi.org/10.3847/1538-4365/ab06fc

Hoy, C., & Raymond, V. (2021). PESUMMARY: The code agnostic Parameter Estimation Summary page builder. *SoftwareX*, *15*, 100765. https://doi.org/10.1016/j.softx.2021.100765

Khan, S., Ohme, F., Chatziioannou, K., & Hannam, M. (2020). Including higher order multipoles in gravitational-wave models for precessing binary black holes. *Phys. Rev. D*, *101*(2), 024056. https://doi.org/10.1103/PhysRevD.101.024056

Lange, J., O'Shaughnessy, R., & Rizzo, M. (2018). Rapid and accurate parameter inference for coalescing, precessing compact binaries. *arXiv e-Prints*. http://arxiv.org/abs/1805.10457

Ossokine, S., & others. (2020). Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation. *Phys. Rev. D*, *102*(4), 044055. https://doi.org/10.1103/PhysRevD.102.044055

Pratten, G., & others. (2021). Computationally efficient models for the dominant and sub-dominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, *103*(10), 104056. https://doi.org/10.1103/PhysRevD.103.104056

The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adhikari, N., Adhikari, R. X., Adya, V. B., & al., et. (2021). GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *arXiv e-Prints*, arXiv:2108.01045. http://arxiv.org/abs/2108.01045

The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adhikari, N., Adhikari, R. X., & al., et. (2021). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv e-Prints*, arXiv:2111.03606. http://arxiv.org/abs/2111.03606

Veitch, J., & others. (2015). Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, *91*(4), 042003. https://doi.org/10.1103/PhysRevD.91.042003

Williams, D., Macleod, D., Vajpeyi, A., & Clark, J. (2021). *Transientlunatic/asimov*. https://doi.org/10.5281/zenodo.4024432