Applicability Study of the PRIMAD Model to LIGO Gravitational Wave Search Workflows

Dylan Chapp
dchapp@udel.edu
University of Delaware

Danny Rorabaugh
dror@utk.edu
University of Tennessee, Knoxville

Duncan A. Brown
dabrown@syr.edu
Syracuse University

Ewa Deelman
deelman@isi.edu
University of Southern California

Karan Vahi
vahi@isi.edu
University of Southern California

Von Welch
vwelch@iu.edu
Indiana University, Bloomington

Michela Taufer
taufer@gmail.com
University of Tennessee, Knoxville

ABSTRACT

The PRIMAD model with its six components (i.e., Platform, Research Objective, Implementation, Methods, Actors, and Data) provides an abstract taxonomy to represent computational experiments and promote reproducibility by design. In this paper, we employ a post-hoc assessment of the model applicability to a set of Laser Interferometer Gravitational-Wave Observatory (LIGO) workflows from literature sources (i.e., published papers). Our work outlines potential advantages and limitations of the model in terms of its levels of abstraction and means of application.

KEYWORDS

Reproducibility; LIGO; Workflows; PRIMAD

ACM Reference Format:
Dylan Chapp, Danny Rorabaugh, Duncan A. Brown, Ewa Deelman, Karan Vahi, Von Welch, and Michela Taufer. 2019. Applicability Study of the PRIMAD Model to LIGO Gravitational Wave Search Workflows. In *2nd International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS'19), June 24, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3322790.3330591

1 INTRODUCTION

The ability of the scientific community to incrementally build on experimental results depends strongly on the ability to trust that those results are not accidental or transient, but rather that they can be reproduced to an acceptably high degree of similarity by subsequent experiments. This notion of reproducibility is magnified both in importance and difficulty in the setting of computational science workflows [39, 40]. An increasingly large fraction of scientific endeavors depend on or incorporate computational elements, which in turn opens the door to reproducibility challenges associated with the implementation of those computational elements.

In order to reason about and assess reproducibility in the computational context, the PRIMAD model [21] was proposed. PRIMAD breaks reproducibility into six named components (i.e., Platform, Research objective, Implementation, Methods, Actors, and Data), each of which represents an element of a computational experiment where reproducibility can be enforced by design, or conversely where a lack of such design can allow irreproducibility to seep in and corrode the overall integrity of the experiment.

To evaluate the efficacy of PRIMAD as a tool for characterizing the reproducibility of real-world computational science workflows, we examine computational workflows used to detect gravitational waves using data from the Laser Interferometer Gravitational-Wave Observatory (LIGO) [1] and the Virgo Observatory [10]. These computational workflows are designed to detect various astronomical events, including binary black hole mergers [3–7] and binary neutron star mergers [8].

There are three factors that make the gravitational-wave search workflows particularly appropriate for our post-hoc study through the lens of PRIMAD: (1) LIGO and Virgo have reached a mature status with findings that have been recognized by the scientific community at large; (2) gravitational-wave search workflows support high impact scientific findings (i.e., empirical confirmation of the existence of gravitational waves) that are subject to the highest levels of scrutiny from the broader scientific community; and (3) the workflows consume a large amount of data and have high internal complexity, leading to reproducibility challenges in terms of the implementation and data management components of PRIMAD.

We tackle the study of gravitational-wave search workflows in a post-hoc fashion using literature sources (i.e., published papers), rather than from the runtime point of view. The study mimics the effect of scientists in replicating the work done by others as a starting point for new research. Specifically, we study two gravitational-wave searches presented by the LIGO Scientific Collaboration and the Virgo Collaboration, and a third search presented by an independent group of scientists[1]. These are: (1) the O1 Binary Catalogue [9];

P-RECS'19, June 24, 2019, Phoenix, AZ, USA

ACM ISBN 978-1-4503-6756-1/19/06…$15.00
https://doi.org/10.1145/3322790.3330591

[1] We note that this independent group contained former LIGO Collaboration members. However, they only had access to publicly available data, code, and information in their analysis.

(2) the O2 Binary Catalogue [23]; and (3) the Open Gravitational Wave Catalogue (1-OGC) [29]. The findings are reported in this paper and organized in terms of the six components of PRIMAD.

2 THE PRIMAD MODEL

Large-scale scientific applications are composed of complex workflows, which in turn are composed of and execute a series of experimental, computational, and data manipulation steps in one or multiple scientific domains, with at least one computational element. Research teams executes such workflows to gain insights that (1) confirm or disprove a hypothesis or (2) discover novel behavior or phenomena.

The PRIMAD Model [21] is a powerful method to describe these workflows in terms of this 6-component space:

P = Platform / execution environment / context
R = Research objectives / goals
I = Implementation / code / source-code
M = Methods / algorithms
A = Actors / persons
D = Data

The model integrates multiple degrees of abstraction and is moldable to fit the characteristics of different scientific studies across domains.

A *platform* can be a simple execution environment such as a virtual machine or container; a more sophisticated combination of high-end clusters for computing and data processing; or a combination of remote sensing devices plugged into systems for collecting data, processing the data in preparation for analysis, analyzing the data, and displaying the results possibly through data visualization.

A *research objective* includes a specification of acceptable result variability provided by the scientist. If the scientist requires bitwise identical results (e.g., for regulatory reasons or out of technical necessity as in some techniques from mathematical physics and experimental mathematics [12]), then this specification must be incorporated into the research objectives. Alternatively, if the scientist requires looser tolerances (e.g., convergence to a certain energy threshold as in ensemble simulations of protein folding trajectories [34]), this specification should be expressed as part of the research objective.

The implementation can comprise a single executable or source-code (e.g., from benchmark applications such as the CORAL procurement benchmarks [45], to production-grade applications such as ALE3D [24]), or multiple software artifacts (e.g., solver packages such as HYPRE [20]) that support one or more research objectives.

Methods are broadly defined by individual computational algorithms, heuristic techniques, step-by-step procedures, or combinations thereof.

Actors are those who perform the scientific study (e.g., scientists, users, technicians). They may have designed the overall workflow or individual parts of it, or simply use it.

Data refers to all the data that are ingested by the initial stage of a workflow and those data objects that intermediate stages of the workflow generate.

3 THEORETICAL SCENARIOS EXHIBITING REPRODUCIBILITY CONCEPTS

Changes in the six components of the model generate different scenarios that exhibit different reproducibility behaviors [16, 19]. We envision six scenarios of interest (i.e., scenarios that are plausible for real-world scientific workflows). These scenarios are summarized in Table 1. Note that this table is not meant to be an exhaustive list.

	Platform (P)	Research (R)	Implementation (I)	Methods (M)	Actors (A)	Data (D)
Scenario I	same	same	same	same	same	same
Scenario II	same or similar	same	same	same	different	same
Scenario III	same or similar	same	same or similar	same or similar	different	same
Scenario IV	same or similar	same	same or similar	similar or different	different	different
Scenario V	same or similar	same	different	similar or different	similar or different	similar or different
Scenario VI	different	same	different	different	different	different

Table 1: Overview of six scenarios exhibiting relevant reproducibility outcomes. Concepts like similar mean some overlapping while different implies complete disjoint component features or items.

The first scenario is one in which the six PRIMAD components are all the same. In other words, the workflow is replayed using the same data and the same computational (software and hardware) environment. The same team of actors is able to produce results that confirm or disprove the hypothesis previously tested with the same workflow. Actors can repeat a measure over and over again, within the expected precision or threshold of acceptable variability. Challenges associated to this scenario are: (1) the platform availability can be an issue when dealing with workflows running on, for example, leadership class computing clusters; (2) documentation deficiencies can be an issue when dealing with complex workflows, for example, those concerning multiscale modeling simulations; and (3) run-to-run variability can occur, even when all six factors of the PRIMAD model are held fixed, due to hardware characteristics, platform concurrency, nondeterministic application behavior, and other factors. Specifically, aspects of the chosen methods or implementation may be inherently nondeterministic, as in the case of dynamic multithreading [43].

The second scenario is one whose actors change, (e.g., a different team in the same or a different institution) and while the platform may be different, such differences shall be limited (e.g., different operating system version, upgraded library). In other words, by using the same data and the same or similar computational (software and hardware) platform, a different team of actors is able to produce results that confirm or disprove the hypothesis previously tested with the same workflow. In addition to the challenges of Scenario 1, this scenario is vulnerable to improper documentation dissemination, for example, sharing documentation with incomplete or

ambiguous descriptions. This scenario and its challenges match such efforts as the Student Cluster Competition [22].

The third scenario explores the impact of varying, to different degrees, all PRIMAD components of the model except the research objectives and the data. By using the same data (e.g., the same initial conditions such as pressure, temperature, and volume in a molecular dynamic simulation) but allowing limited variability of implementations and platforms or both, along with published material and related artifacts, an independent team is able to produce result that confirm or disprove the same hypothesis presented in the published material. The challenges associated with this, and following scenarios, depend on the degree of variation in the individual components of the PRIMAD model that we allow.

The fourth scenario is a scenario that is less constrained than the previous one. It explores the impact of varying all PRIMAD components except the research objectives. This scenario incorporates the additional challenge of collecting new data according to the data generation description in one or multiple original publications. The study can be performed by using the same or similar implementations and platforms. In other words, minor variations in implementation and platform are acceptable in this scenario. Along with the published material and related artifacts, an independent team is able to produce results that confirm or disprove the same hypothesis.

The fifth scenario studies the ability of a team of actors to reproduce the research of another team using a different implementation (e.g., in molecular dynamics simulations using AMBER rather than CHARMM). They address the same research objectives, but the level of variability in workflow outcomes depends upon the changes in implementation.

The sixth scenario tackles the study of a scientific application for which the workflow has a substantial change of the platform and implementation. Specifically, this scenario deals with a major technological transformation (e.g., from multiprocessor to accelerated platforms). All components of the model space in which the workflow is executed, with the exception of the research objective, change sufficiently and thus these variations cannot be ignored.

4 EMPIRICAL SCENARIOS EXHIBITING REPRODUCIBILITY CONCEPTS IN THE LIGO PROJECT

Our post-hoc study of the PRIMAD applicability to gravitational-wave search workflows falls into Scenario V. When we use the term workflow, we refer to the analysis of a specific set of data using a specific search pipeline. In practice, this may involve running several identically-configured workflows on smaller data sets to produce a single search result. Specifically, we target the gravitational wave search in three literature sources: (1) the LIGO-Virgo O1 Binary Catalogue [9]; (2) the LIGO-Virgo O2 Binary Catalogue GWTC-1 [23]; and (3) the Open Gravitational Wave Catalogue (1-OGC) [29], produced independently of the LIGO-Virgo Collaborations. The two LIGO-Virgo Binary Catalogue papers [9, 23] describe two matched filtering workflows built on two implementations or source codes (with PyCBC [2, 28, 31, 42] and GstLAL [2, 25, 26, 37] data analysis suites, respectively). The LIGO-Virgo O1 Binary Catalogue paper [9] and its constituent workflows use the dataset from Observing Run 1 of the LIGO detectors (O1). The LIGO-Virgo O2 Binary Catalogue paper [23] and its constituent analyses use the dataset from Observing Run 2 of the LIGO detectors (O2) as well as data from the VIRGO detector. The 1-OGC paper [29, 30] describes a run of the PyCBC matched filtering workflow on the public LIGO O1 data using an updated pipeline configuration compared to the analysis in the LIGO-Virgo O1 Binary Catalogue. The relationships between the workflows and their parent publications are summarized in the first two rows of Table 2 together with the six components of the PRIMAD model.

PRIMAD Component	O1 Binary Catalog [9]		O2 Binary Catalog [23]		1-OCG [29]
	O1BC-PyCBC	O1BC-GstLAL	O2BC-PyCBC	O2BC-GstLAL	1-OGC
(P)latform	Proprietary		Proprietary		Public
(R)esearch Objective	Discover BBH Mergers		• Discover BBH Mergers • Re-analyze O2 with improved data quality • Re-analyze O1 with improved pipelines		• Re-analyze O1 with improved pipelines • Demonstrate reproducibility of O1 result on public data • Produce and release sub-threshold events
(I)mplementation	See implementation comparison table				
(M)ethod/ Algorithm	Matched filtering				
(A)ctors	LIGO collaboration authors				OGC authors
(D)ata	See data comparison table				

Table 2: Overview of LIGO analyses considered in our case study in terms of the 6 PRIMAD model components.

Implementation Component	O1BC-PyCBC	O2BC-PyCBC	O1BC-GstLAL	O2BC-GstLAL	1-OGC
Template bank	TWB-O1BC See [15, 36, 39]	TWB-O2-PyCBC See [18]	TWB-O1BC See [15, 36, 39]	TWB-O2-GstLAL See [28]	TWB-OGC See [14, 32, 33]
Waveform model	Delegated to multiple refs.	SEOBNRv4 + TaylorF2 waveform approximant	Delegated to multiple refs.	SEOBNRv4 + TaylorF2 waveform approximant	SEOBNRv4 + TaylorF2 waveform approximant
Template bank decomposition	Delegated to companion papers or proprietary				15 equally sized sub-banks
Noise-filtering preprocessing	Delegated to [2]	Delegated to [37]	Delegated to [2]	Delegated to [37]	Delegated to [28, 40]
Single-detector SNR threshold/conditions	Delegated to [2]	>5.5	Delegated to [2]	>4 (LIGO) and >3.5 (VIRGO)	>4 and >any other peaks w/in 1s
Ranking statistic	Background using time slide method	Improved version of statistic used in O1-PyCBC	Logarithm of likelihood ratio	Logarithm of likelihood ratio + Monte Carlo See [37]	Background using time slide method

Table 3: Matched filtering implementation across the LIGO data workflows targeted in our study (i.e., O1BC-PyCBC, O2BC-PyCBC, O1BC-GstLAL, O2BC-GstLAL, 1-OGC).

In the table, the artifacts are executed on proprietary or public *platforms*. We consider proprietary platforms to be those platforms that are only the actors of the workflow have access to. In contrast we consider platforms to be public when they are accessible to a broader set of potential actors (e.g., the Open Science Grid (OSG) [35, 38]).

The research objective of the workflows is to detect binary mergers using the data gathered by the LIGO and Virgo detectors. The workflows may also re-analyze previous data with improved methods and tuning to re-assess the significance of previous events, or discover previously undetected events; this is considered similar enough to the core research objective to be considered the same within the context of PRIMAD.

Data Component	O1BC-PyCBC	O2BC-PyCBC	O1BC-GstLAL	O2BC-GstLAL	1-OGC
Observing Runs	O1	O1+O2	O1	O1+O2+Virgo	O1 (public subset only)
# days available for coincident analysis	51.5	118	51.5	118 (15 w/ Virgo)	51.5
Metadata used to decide keep/discard	Proprietary (data quality flags for vetoing events)				Public
# days after quality filtering	48.6	Delegated to refs.	48.6	Delegated to refs.	48.1
# days after duration filtering	46.1		48.3		

Table 4: Data details across the targeted LIGO workflows targeted in our study (i.e., O1BC-PyCBC, O2BC-PyCBC, O1BC-GstLAL, O2BC-GstLAL, 1-OGC).

The descriptions of matched filtering workflows provide numerous details in terms of implementation choices (e.g., choices of signal-to-noise ratio (SNR) thresholds for detecting significant matches) and data management choices (e.g., choices for how detector instrument noise is filtered and mitigated). The abundance of details concerning implementation and data variations across the matched filtering studies are summarized in Table 3 and Table 4 respectively.

Descriptions of the search pipeline implementation are listed in the leftmost column of Table 3. These sub-components define the particular way that a workflow (e.g., the O1BC-PyCBC workflow) implements matched filtering (i.e., the method component). Specifically, we have highlighted the following implementation sub-components: the template bank used, the waveform model used to generate the template bank, the policy used to decompose the template bank into sub-banks, any additional noise reduction or filtering steps performed, the signal-to-noise ratio threshold for selecting candidate matches (i.e., events that could potentially be binary black holes), and the ranking statistic used rank candidate events.

For the template bank component, we abbreviate the banks used for each workflow. TWB-OGC refers to the template bank used in the 1-OGC search workflows, the properties of which are defined in [14, 32, 33]. TWB-O1BC refers to the common template bank used in both the O1BC-PyCBC and O1BC-GstLAL workflows, the properties of which are defined in [15, 36, 41]. TWB-O2BC-PyCBC refers to the template bank used for the O2 Binary Catalogue PyCBC workflow, the properties of which are defined in [17]. TWB-O2BC-GstLAL refers to the template bank used for the O2 Binary Catalogue GstLAL workflow, the properties of which are defined in [26].

We find that for some components of the workflow, details are predominantly delegated to one or more cited publications. In other cases, configuration of the workflow is only available by examining the workflow's source code or the configuration files used in a specific analysis. We also find that some workflow configuration files are public (e.g. [27, 30]), but others require proprietary access. Even for configuration and software repositories that are public, it can be difficult to determine which versions were used for a particular analysis. Similarly for Table 4, we describe sub-components of the Data PRIMAD component in the leftmost column. The sub-components we describe include: the source of the raw input data (i.e., which observing runs the data comes from), the number of coincident days of data available (i.e., data from when both LIGO detectors were operating), the choice of metadata used to filter data prior to the matched filtering search, the number of days of observation data remaining after filtering for data quality, and the number of days of observation data remaining after filtering insufficiently long-lasting events (i.e., filtering based on event duration). These details are largely provided across the chosen workflows, unless those details are proprietary.

Across all of the workflows described in the LIGO papers used for our study, the high-level method that is implemented is matched filtering [11], that is, the procedure of matching a stream of data from the LIGO detectors against template waveforms derived from numerical general relativity. Matches against these templates may indicate the presence of astronomical events such as binary black hole mergers, and thus are evaluated for statistical significance.

5 DISCUSSION AND CONCLUSIONS

We conclude with a reflection on PRIMAD, beginning with general benefits and challenges of the model itself. Then we evaluate our post-hoc application of PRIMAD to the LIGO workflows, which informs best practices for future runtime applications of PRIMAD in the execution of scientific workflows.

PRIMAD is a general model to guide reproducibility. It helps meet an acute need in the scientific community to ground reproducibility, yet it is inherently abstract due to its applicability across all scientific domains, leading to challenges in establishing a useful level of specificity. When researchers want to share their findings, many research teams (or *actors*) make best efforts at including all relevant information to enable reproducibility. Without standardization, the decisions about what constitutes "relevant information" are inevitably ad-hoc, and may not be uniform from publication to publication or across multiple workflows within a single publication. Thus, PRIMAD offers a framework with which to build sustainable reproducibility in an uniform fashion across scientific domains.

Challenges arise in identifying how specific the components of PRIMAD need to be defined in order to guarantee both consistent applicability across workflows in a specific domain and desired levels of reproducibility. For example, the division between implementation and methods is disputable. Minor adjustments to an algorithm would generally fall into implementation, yet it is hard to determine when changes are substantial enough to call it a new algorithm and thus a change in methods. In other cases, the effects of the actors on reproducibility may be difficult to document. Even within the same research group with consistent leadership, research objectives, and computational environments, changes in team members and shifts in member responsibility can introduce unacknowledged sources of variability. It appears to be a difficult problem to appropriately document the knowledge and experience that is applied to the elements of a workflow. Finally, different datasets may play different roles within a workflow. As a partial remedy, we propose a differentiation between input data (including initial parameters), intermediary data (which depends heavily on

methodology and implementation), and output artifacts (having direct relevance to the research objective). In order to meet these challenges, we recommend that each field of science develop its own domain-appropriate refinement to PRIMAD.

The LIGO workflows demonstrate scientific research performed with a clear desire for reproducibility. Therefore, even though the authors do not apply an explicit reproducibility framework, we were able perform a post-hoc evaluation of the workflows with the PRIMAD model. In particular, we find that out of the six PRIMAD components, the implementation and data components are given the most attention in terms of documented details, as summarized in Table 3 and Table 4. One PRIMAD component not described with the same level of details as the other components is the platform used in the workflows. This may be due to multiple factors such as: (1) use of proprietary platforms; (2) prioritization of implementation or data management details over platform details; and (3) the technical challenge of tracking all the platform components on which a workflow manager deploys jobs. Uncertainty in the reason for incomplete information is one of a few limitations we found in post-hoc application of PRIMAD. Non-uniform descriptions of PRIMAD components and the necessity of citation chasing to fill in details also arise, though are mitigated by documentation efforts of actors within a study and overlap of actors between studies.

Challenges listed above can be mitigated by moving from a post-hoc to a runtime application of PRIMAD from the early stages of scientific experimentation. The PRIMAD components should be periodically referred to as a checklist for assessing the reproducibility of workflows, both as they are developed and during the transition from experimentation to publication. Moreover, rather than hampering scientific progress by imposing requirements inconsistent with a particular scientific workflow, we demonstrated with the various scenarios in Section 4 how such a framework can be a guide to ensure the desired level of reproducibility.

A wide adoption of PRIMAD would be most achievable through involvement of all contributors to the scientific process. That is, tools that researchers use to do computational experiments must: (1) be aware of the PRIMAD components; and (2) provide users with access to details about their experiments, organized in terms of the PRIMAD components, so that those details can be disseminated in publications. In particular, we look at the role of workflow managers in the computational components of a scientific workflow. In the case of the PyCBC matched filtering pipelines, the actors used Pegasus [13, 18] to distribute constituent jobs to participating computing centers [44]. In the long run, workflow managers such as Pegasus can play a key role in tracking, for example, which computational platforms their jobs run on so that future research teams can attempt to reproduce results on similar platforms, and therefore automatically, deploying the PRIMAD model at runtime.

As part of our study of gravitational-wave search pipelines, both from the LIGO collaboration and from others, we have identified the various implementation and data components that need to be identified in order for an outside actor to reproduce the results of these pipelines. We feel this axis of classification can serve as a foundation for describing all (or a large number of) gravitational-wave search pipelines and lowering the barriers for reproducing the results outside of the LIGO collaboration.

ACKNOWLEDGMENTS

The authors thank Almadena Chtchelkanova and the U.S. National Science Foundation for sponsoring the July 25, 2017 workshop that provided the foundation on the PRIMAD model. The authors also thank the workshop attendees: Michael A. Heroux, Sandia National Laboratories; Lorena A. Barba, Ronald Boisvert, National Institute of Standards and Technology; Bruce Childers, University of Pittsburgh; Juliana Freire, New York University; Carlos Maltzahn, University of California, Santa Cruz; Wilf Pinfold, Urban.Systems; Jeff Spies, Center for Open Science. The work was supported by NSF awards OAC-1841399, OAC-1823405, OAC-1823378, and CCF-1841552.

REFERENCES

[1] J. Aasi et al. 2015. Advanced LIGO. *Class. Quant. Grav.* 32 (2015), 074001. https://doi.org/10.1088/0264-9381/32/7/074001

[2] B. P. Abbott et al. 2016. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D* 93, 12 (2016), 122003. https://doi.org/10.1103/PhysRevD.93.122003

[3] B. P. Abbott et al. 2016. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.* 116, 24 (2016), 241103. https://doi.org/10.1103/PhysRevLett.116.241103

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

[5] B. P. Abbott et al. 2017. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.* 118, 22 (2017), 221101. https://doi.org/10.1103/PhysRevLett.118.221101

[6] B. P. Abbott et al. 2017. GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence. *Astrophys. J.* 851, 2 (2017), L35. https://doi.org/10.3847/2041-8213/aa9f0c

[7] B. P. Abbott et al. 2017. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Phys. Rev. Lett.* 119, 14 (2017), 141101. https://doi.org/10.1103/PhysRevLett.119.141101

[8] B. P. Abbott et al. 2017. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.* 119, 16 (2017), 161101. https://doi.org/10.1103/PhysRevLett.119.161101

[9] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. 2016. Binary black hole mergers in the first advanced LIGO observing run. *Phys. Rev. X* 6, 4 (2016), 041015. https://doi.org/10.1103/PhysRevX.6.041015

[10] F. Acernese et al. 2015. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* 32, 2 (2015), 024001. https://doi.org/10.1088/0264-9381/32/2/024001

[11] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. 2012. FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D* 85 (2012), 122006. https://doi.org/10.1103/PhysRevD.85.122006

[12] David H. Bailey, Roberto Barrio, and Jonathan M. Borwein. 2012. High-precision computation: Mathematical physics and dynamics. *Appl. Math. Comput.* 218, 20 (2012), 10106–10121. https://doi.org/10.1016/j.amc.2012.03.087

[13] Duncan A. Brown, Patrick R. Brady, Alexander Dietz, Junwei Cao, Ben Johnson, and John McNabb. 2007. A case study on the use of workflow technologies for scientific analysis: Gravitational wave data analysis. In *Workflows for e-Science*. Springer, London, England, 39–59. https://doi.org/10.1007/978-1-84628-757-2_4

[14] Duncan A. Brown, Ian Harry, Andrew Lundgren, and Alexander H. Nitz. 2012. Detecting binary neutron star systems with spin in advanced gravitational-wave detectors. *Phys. Rev. D* 86, 8 (2012), 084017. https://doi.org/10.1103/PhysRevD.86.084017

[15] Collin Capano, Ian Harry, Stephen Privitera, and Alessandra Buonanno. 2016. Implementing a search for gravitational waves from binary black holes with nonprecessing spin. *Phys. Rev. D* 93, 12 (2016), 124007. https://doi.org/10.1103/PhysRevD.93.124007

[16] Jon F. Claerbout and Martin Karrenbach. 1992. Electronic documents give reproducible research a new meaning. In *SEG Technical Program Expanded Abstracts 1992*. Society of Exploration Geophysicists, Tulsa, OK, 601–604. https://doi.org/10.1190/1.1822162

[17] Tito Dal Canton and Ian Harry. 2017. Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run. (2017), 10 pages. arXiv:gr-qc/1705.01845

[18] Ewa Deelman, Gurmeet Singh, Mei-Hui Su, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Karan Vahi, G. Bruce Berriman, John Good, Anastasia Laity, Joseph C. Jacob, and Daniel S. Katz. 2005. Pegasus: A framework

for mapping complex scientific workflows onto distributed systems. *Scientific Programming* 13, 3 (2005), 219–237. https://doi.org/10.1155/2005/128026

[19] David L. Donoho. 2010. An invitation to reproducible computational research. *Biostatistics* 11, 3 (2010), 385–388. https://doi.org/10.1093/biostatistics/kxq028

[20] Robert D. Falgout and Ulrike Meier Yang. 2002. hypre: A library of high performance preconditioners. In *International Conference on Computational Science (Lecture Notes in Computer Science)*, Sloot P. M. A., Hoekstra A. G., Tan C. J. K., and Dongarra J. J. (Eds.), Vol. 2331. Springer, Berlin, Germany, 632–641. https://doi.org/10.1007/3-540-47789-6_66

[21] Juliana Freire, Norbert Fuhr, and Andreas Rauber. 2016. Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041). *Dagstuhl Reports* 6, 1 (2016), 108–159. https://doi.org/10.4230/DagRep.6.1.108

[22] C. Kris Garrett, Stephen Lien Harrell, and Michael A. Heroux. 2018. Special Issue on SCC'17 Reproducibility Initiative. *Parallel Comput.* 79 (Nov. 2018), 48–49. https://doi.org/10.1016/j.parco.2018.10.001

[23] LIGO Scientific Collaboration, Virgo Collaboration, et al. 2018. GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. (2018), 48 pages. arXiv:astro-ph.HE/1811.12907

[24] M. A. McClelland, E. A. Glascoe, A. L. Nichols, S. P. Schofield, and H. K. Springer. 2014. *ALE3D simulation of incompressible flow, heat transfer, and chemical decomposition of Comp B in slow cookoff experiments.* Technical Report. Lawrence Livermore National Laboratory (LLNL).

[25] Cody Messick, Kent Blackburn, Patrick Brady, Patrick Brockill, Kipp Cannon, Romain Cariou, Sarah Caudill, Sydney J. Chamberlin, Jolien D. E. Creighton, Ryan Everett, et al. 2017. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D* 95, 4 (2017), 042001.

[26] Debnandini Mukherjee et al. 2018. The GstLAL template bank for spinning compact binary mergers in the second observation run of Advanced LIGO and Virgo. (2018). arXiv:astro-ph.IM/1812.05121

[27] Alex Nitz et al. 2018. PyCBC Configuration File Repository. https://github.com/gwastro/pycbc-config

[28] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, John Willis, Tito Dal Canton, Larne Pekowsky, Thomas Dent, Andrew R. Williamson, Collin Capano, Soumi De, Miriam Cabero, Bernd Machenschalk, Prayush Kumar, Steven Reyes, Thomas Massinger, Amber Lenon, Stephen Fairhurst, Alex Nielsen, shasvath, Francesco Pannarale, Leo Singer, Duncan Macleod, Stanislav Babak, Hunter Gabbard, John Veitch, Lorena Magaña Zertuche, Peter Couvares, and Brian Bockelman. 2017. PyCBC: O2 Production Release 20. https://doi.org/10.5281/zenodo.883086

[29] Alexander H. Nitz, Collin Capano, Alex B. Nielsen, Steven Reyes, Rebecca White, Duncan A. Brown, and Badri Krishnan. 2019. 1-OGC: The first open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO data. *The Astrophysical Journal* 872, 2 (2019), 195. https://doi.org/10.3847/1538-4357/ab0108

[30] Alexander H. Nitz, Collin Capano, Alex B. Nielsen, Steven Reyes, Rebecca White, Duncan A. Brown, and Badri Krishnan. 2019. 1-OGC: The first open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO data: Workflow Configuration Files. https://github.com/gwastro/1-ogc/tree/master/workflow/configuration

[31] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. 2017. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.* 849, 2 (2017), 118. https://doi.org/10.3847/1538-4357/aa8f50

[32] Benjamin J. Owen. 1996. Search templates for gravitational waves from inspiraling binaries: Choice of template spacing. *Phys. Rev. D* 53, 12 (1996), 6749. https://doi.org/10.1103/PhysRevD.53.6749

[33] Benjamin J. Owen and Bangalore Suryanarayana Sathyaprakash. 1999. Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement. *Phys. Rev. D* 60, 2 (1999), 022002. https://doi.org/10.1103/PhysRevD.60.022002

[34] Stefano Piana, John L. Klepeis, and David E. Shaw. 2014. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current opinion in structural biology* 24 (2014), 98–105. https://doi.org/10.1016/j.sbi.2013.12.006

[35] Ruth Pordes et al. 2007. The Open Science Grid. *J. Phys. Conf. Ser.* 78 (2007), 012057. https://doi.org/10.1088/1742-6596/78/1/012057

[36] Michael Pürrer. 2016. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass ratios and spins. *Phys. Rev. D* 93, 6 (2016), 064041. https://doi.org/10.1103/PhysRevD.93.064041

[37] Surabhi Sachdev et al. 2019. The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs. (2019). arXiv:gr-qc/1901.08580

[38] Igor Sfiligoi, Daniel C. Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwrin. 2009. The pilot way to Grid resources using glideinWMS. *WRI World Congress* 2 (2009), 428–432. https://doi.org/10.1109/CSIE.2009.950

[39] Victoria. Stodden and Matthew S. Krafczyk. 2018. Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context. (2018).

[40] Victoria Stodden, Marcia McNutt, David H. Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A. Heroux, John P.A. Ioannidis, and Michela Taufer. 2016. Enhancing reproducibility for computational methods. *Science* 354, 6317 (2016), 1240–1241. https://doi.org/10.1126/science.aah6168

[41] Andrea Taracchini, Alessandra Buonanno, Yi Pan, Tanja Hinderer, Michael Boyle, Daniel A. Hemberger, Lawrence E. Kidder, Geoffrey Lovelace, Abdul H. Mroué, Harald P. Pfeiffer, et al. 2014. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev. D* 89, 6 (2014), 061502. https://doi.org/10.1103/PhysRevD.89.061502

[42] Samantha A. Usman et al. 2016. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.* 33, 21 (2016), 215004. https://doi.org/10.1088/0264-9381/33/21/215004

[43] Robert Utterback, Kunal Agrawal, I. Lee, Milind Kulkarni, et al. 2017. Processor-oblivious record and replay, In PPoPP '17. *ACM SIGPLAN Notices* 52, 8, 145–161. https://doi.org/10.1145/3155284.3018764

[44] Derek Weitzel, Brian Bockelman, Duncan A. Brown, Peter Couvares, Frank Würthwein, and Edgar Fajardo Hernandez. 2017. Data Access for LIGO on the OSG. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact.* ACM, New York, NY, 24. https://doi.org/10.1145/3093338.3093363

[45] Xingfu Wu and Valerie Taylor. 2015. Power and performance characteristics of CORAL Scalable Science Benchmarks on BlueGene/Q Mira. In *2015 Sixth International Green and Sustainable Computing Conference (IGSC).* IEEE Computer Society, Washington, DC, 6. https://doi.org/10.1109/IGCC.2015.7393681