Reproducibility of the First Image of a Black Hole in the Galaxy M87 From the Event Horizon Telescope Collaboration

Ria Patel, Brandan Roachell, Silvina Caino-Lores, Ross Ketron, Jacob Leonard, and Nigel Tan, University of Tennessee, Knoxville, Knoxville, TN, 37996, USA
Karan Vahi, University of Southern California, Los Angeles, CA, 90007, USA
Duncan A. Brown, Syracuse University, Syracuse, NY, 13244, USA
Ewa Deelman, University of Southern California, Los Angeles, CA, 90007, USA
Michela Taufer, University of Tennessee, Knoxville, Knoxville, TN, 37996, USA

This article presents an interdisciplinary effort to develop and share sustainable knowledge necessary to analyze, understand, and use published scientific results to advance reproducibility in multimessenger astrophysics. Specifically, we target the breakthrough work associated with generating the first image of a black hole, called M87. The Event Horizon Telescope (EHT) Collaboration computed the image. Based on the artifacts made available by the EHT, we deliver documentation, code, and a computational environment to reproduce the first image of a black hole. Our deliverables support discovery in multimessenger astrophysics by providing all of the necessary tools for generalizing methods and findings from the EHT use case.

Challenges encountered during the reproducibility of EHT results are reported. Our effort results in an open source, containerized software package that enables the public to reproduce the first image of a black hole in the galaxy M87.

Developing reproducible analyses is a challenging aspect of scientific research. Few real-world studies have been performed to guide the necessary processes and products, especially in domains relying on scientific computing. The availability of data, software, platforms, and documentation limits their reproducibility. Consequently, despite a group’s best efforts, other scientists attempting to reproduce an analysis may find the necessary information incomplete.

We present an interdisciplinary effort to develop and share sustainable knowledge necessary to understand, reproduce, and reuse the published scientific results of the Event Horizon Telescope (EHT) project’s analysis of the black hole in the center of the M87 galaxy. Unlike our previous reproduction of Advanced LIGO’s observations, none of the authors of this article was involved in the original EHT analysis. Thus, our work builds exclusively on several articles describing the EHT project workflow, data, and software that are available online. Each EHT paper presents specific aspects of scientific discovery. Several groups have independently validated the EHT results. This article is not another independent analysis of the EHT results but describes the complete software, environment, and documentation needed to reproduce the published results. The reproducibility effort follows rigorous
reproducibility directions and expands preliminary work presented in a poster. The rebuilt EHT software pipelines are made available to scientists for reuse and further scientific discovery.

As part of our contributions, we investigate the availability and integrity of the data used to recreate the images of the M87 black hole. We model the image processing workflow and study its limitations regarding software availability, dependencies, configuration, portability, and documentation. We rebuild the workflow’s software stack to reproduce the published images; we use the software stack to analyze discrepancies between the original and reproduced results. We document each step in this process, starting from a systematic assessment of data, software, and documentation availability. We deliver a collection of fully documented containers for data validation and image reconstruction. Finally, we compile guidelines to increase the reproducibility of computational workflows in scientific projects. Our work enhances the reproducibility and reach of scientific projects, such as the EHT project, and facilitates the engagement of the overall scientific community, including postdocs and students, regardless of the domain.

M87 EHT

The EHT project uses very-long-baseline interferometry (VLBI) to link together eight radio telescopes around the world to study the immediate environment of a black hole with angular resolution comparable to the size of the black hole itself. In April 2019, the EHT Collaboration published measurements of the properties of the central radio source in M87, including the first direct image of a black hole. The results, which received worldwide attention, revealed, for the first time, a bright ring formed as light bends in the intense gravity around a black hole in the galaxy M87. The black hole is 6.5 billion times larger than the sun.

The EHT project provides links to its calibrated data published in CyVerse Data Commons, a publication describing the project’s data processing and calibration, a link to the software used in the imaging workflow, and a publication describing the imaging workflow. The EHT Collaboration released data products and software, hosting them on third-party repositories. This is a common approach for many National Science Foundation-funded projects ranging in size from individual investigators to international collaborations.

CHARACTERIZATION OF THE EHT WORKFLOW

The EHT workflow comprises three key components: data collection, processing, and image building (see Figure 1).

Data Collection

Eight telescopes in the EHT network collect radio interferometry data on certain days at certain times that have permissible weather conditions for all sites, allowing the gathering of data from multiple angles and effectively turning Earth into one single giant telescope. The EHT data used to generate the first M87 black hole images consist of spatiotemporal data of visibility amplitudes collected over five days in 2017.
Collected raw data for each day contain high and low telescope frequencies.

Raw Data Processing

Raw data from the different observatories is first pieced together by using Earth’s geometry and clock/delay model to obtain a common time reference, and the pairwise correlation coefficients are computed. Fringe-fitting is performed using the EHT-HOPS Pipeline for Millimeter VLBI Data Reduction. Data undergo a priori calibration and network calibration in the postprocessing stage of the EHT-HOPS pipeline to create .uvfits files. Then, the data are reduced to a manageable size for source imaging and model fitting; data are fringe-fitted, calibrated a priori, and network calibrated. The processed data are stored in the First M87 EHT Results\textsuperscript{5} data repository in .csv, .txt, and .uvfits formats and are available to the community. The EHT Data Products page has been updated since then with the raw data.\textsuperscript{14} At the time of this study, the raw data were not open access, and the processing scripts were not open source. Thus, we use the processed data for this study.

Image Building

To reduce biases and increase trust in results, the EHT Collaboration uses three independently designed pipelines to generate the black hole images. They are the Difmap M87 Stokes I Imaging Pipeline (DIFMAP),\textsuperscript{15} the EHT-Imaging M87 Stokes I Imaging Pipeline (EHT-Imaging),\textsuperscript{16} and the Sparse Modeling Imaging Library for Interferometry M87 Stokes I Imaging Pipeline (SMILI).\textsuperscript{17} Each pipeline uses different methods, algorithms, and software libraries but uses the same input data. While the code for each pipeline is available as open source software, the repositories do not contain all of the scripts for image postprocessing and generation. Providing documentation for scientific software is challenging. We find that documentation for packaging, installing, and running the pipelines can be incomplete or unavailable for certain parts of the analysis.

Table 1 lists the available, unavailable, and incomplete data, scripts, and pipelines used by the EHT workflow and shared with the community before our reproducibility study. To succeed in our effort, we generated and made available the missing components.

### TABLE 1. Availability of data, scripts, and pipelines before our reproducibility study. Available and incomplete components are linked to the paper or article presenting them; missing components are marked as unavailable.

| Data                     | Raw data          | Available\textsuperscript{14} |
|--------------------------|-------------------|-------------------------------|
| Processed (calibrated) data | Available\textsuperscript{5} |                      |

| Scripts                  | Raw data processing | Unavailable |
|--------------------------|---------------------|-------------|
| Processed data validation | Unavailable         |             |
| Image post-processing    | Unavailable         |             |
| Figure generation        | Unavailable         |             |

| Pipelines               | DIFMAP              | EHT-Imaging       | SMILI |
|-------------------------|---------------------|-------------------|-------|
| Code                    | Available\textsuperscript{15} | Available\textsuperscript{16} | Available\textsuperscript{17} |
| Packaging               | Unavailable         | Unavailable       | Unavailable |
| Installation manual     | Incomplete\textsuperscript{15} | Incomplete\textsuperscript{16} | Incomplete\textsuperscript{17} |
| User manual             | Incomplete\textsuperscript{15} | Incomplete\textsuperscript{16} | Incomplete\textsuperscript{17} |

### VALIDATING THE DATA INTEGRITY

Data integrity validation is a vital aspect of any work reproducing scientific results: the data used to generate the original EHT images should match the data made available to the community. Data integrity is often considered secondary but can compromise any reproducibility effort, as previously demonstrated by Brown et al.\textsuperscript{2} Figure 1 in the article by the EHT Collaboration\textsuperscript{3} characterizes the original data in terms of telescope baselines (i.e., u-v coverages). Scripts to compare the properties of the original data with available data were not available. We generated the missing Python scripts and integrated them into a Jupyter notebook using standard Python modules such as matplotlib, pandas, and numpy.

Figure 2 shows the comparison between the properties of the original data used in Figure 1 in the EHT Collaboration article\textsuperscript{3} [the set of subfigures in Figure 2(a)] and the reproduced properties using the available data [the set of subfigures in Figure 2(b)]. The top left plots in the two sets represent the intrasite EHT interferometer baselines (short baselines). The top right plots represent the aggregate baseline coverage of the EHT array for all four days observed. The bottom plots...
show the short and long baseline coverage observed by each telescope set at high and low frequencies each day. Qualitatively, we can assess the integrity of the data that we input to the three pipelines (i.e., DIFMAP, EHT-imaging, and SMILI). The only difference is the incomplete left plot in Figure 2(b) because the analysis is based on both the available processed EHT data and the unavailable intra-ALMA data from the Atacama Large Millimeter/Submillimeter Array (ALMA). This external dataset is not included in the calibrated EHT Data Products\textsuperscript{5}; based on communications with the EHT Collaboration, the data are not needed for the pipelines to reproduce the black hole images.

**REBUILDING THE EHT SOFTWARE STACK**

The three EHT pipelines that are part of image building can be modeled in terms of their functional modules (Figures 3(a), 4(a), and 5(a)). Each pipeline comprises...
a parameter definition module for users to establish workflow-specific behavior as well as data preparation and data precalibration modules to preprocess the input files that are fed to the core of each pipeline. A module performing the image reconstruction cycles runs the image reconstruction algorithm; note how each pipeline uses a different number of cycles. The output of each pipeline includes a final image and statistics module that is used for qualitative and quantitative analysis of the reconstructed results, respectively. In SMILI, the first two modules are inverted, and a image evaluation module for data visualization is available at the end of the pipeline.

Although the three pipelines share similar high-level steps, each has additional steps, dependencies, and implementation. Figures 3(b), 4(b), and 5(b) show each pipeline’s dependencies and software components.

**DIFMAP**

DIFMAP (Figure 3) is written in C and uses the CLEAN algorithm for image reconstruction involving iterative deconvolution, paired with a technique called “difference mapping.” EHT’s DIFMAP script takes a file containing observation data; a mask (set of cleaning windows) file that defines areas of interest for the algorithm to iterate upon; and five command-line arguments, which have been provided in the EHT repository. After loading this file, the script initializes values, reads the file specifying the mask, and begins the precalibration phase, which involves its first cleaning and phase of self-calibration. Afterward, the image undergoes 20 rounds of amplitude self-calibrations and cleanings, which is when image reconstruction occurs.

**EHT-Imaging**

EHT-Imaging (Figure 4) uses the regularized maximum likelihood (RML) method of image reconstruction and relies heavily on the eht-imaging Python module (EHTIM) to complete its processes. The EHTIM module defines numerous classes to allow the loading, simulation, and manipulation of VLBI data. By leveraging the classes in this module, the EHT-Imaging workflow loads the low- and high-band data files of a single day’s observations into a data object. It performs various data preparation and precalibration steps. The workflow then moves to an imaging cycle with four iterations. Each successive iteration relies directly on the image generated in the previous iteration. After four iterations, the final image is output. The pipeline also allows for optional results, including the final image.

**Figure 3.** Detailed view of the DIFMAP pipeline and its implementation.

**Figure 4.** Detailed view of the EHT-Imaging pipeline and its implementation. EHTIM: eht-imaging Python module; RML: regularized maximum likelihood.
and an image summary file containing various imaging parameters and data related to the imaging process.

SMILI

SMILI (Figure 5) is also written in Python and uses RML like EHT-Imaging. Before imaging, SMILI uses the EHTIM module to use data sets precalibrated consistently with the other workflows. After the precalibration stage, the software generates data tables for the final imaging process. Reconstruction of an image begins with a circular Gaussian with successive iterations relying on the image generated from the previous iteration. There are four stages of iterations, with each stage performing three imaging cycles. Once completed, the software outputs the final image and packages the input, precalibrated, and self-calibrated data files for traceability.

Note that each pipeline has its own GitHub repository.15–17 Compiling each pipeline’s original code from the three EHT repositories resulted in several errors. For example, on a Power9 system, we missed dependencies and had to remove optimization compilation flags from the installation script to generate the executable code successfully. We solved dependencies manually by editing complex scripts; we used Spack, Anaconda, and Pip to install the latest stable version of each necessary library. In general, none of the three pipeline codes include a comprehensive list of required software dependencies and libraries used or their performances. Once the compilation was completed, we experienced runtime errors with EHT-Imaging and SMILI that we solved by correcting syntax issues in a part of the Python code. We could not find documentation on how to transform the gray-scale output of DIFMAP and SMILI into the colored and formatted images from Figure 11 in the EHT collaboration article.3 We solved this issue by utilizing the EHTIM module to postprocess gray-scale output. In rebuilding the EHT software stack, we documented the software packages used, their dependencies, the compilation requirements, and the execution processes for all three pipelines, completing the unavailable or incomplete components in Figure 1.

PACKAGING AND DISTRIBUTION

To support the portability of the EHT workflow across different platforms, we created a collection of four Docker containers that allows users to reproduce two critical results from the EHT project: the characterization of available data (i.e., Figure 1 in the EHT Collaboration article3) and the final EHT images of the black hole in Figure 11 in the EHT Collaboration article3). A first container hosts the entire setting to reproduce the data integrity validation; it includes the data tarball from the EHT Data Product page and our Bash, Python, and Docker scripts. We developed these scripts to automate the installation and configuration of the environment in an easily accessible and portable way. For users to be completely satisfied with the data integrity validation, we have incorporated a spare tarball within the container so that users can perform the md5sum program on it to compare with the md5sum of the data from the EHT Data Products page. If both results match, then users know that the data in the Data Products page have not been modified in any way, and they can move on with the validation by running the Python scripts to reproduce the images of the black hole. The other three containers reproduce the final EHT images of the black hole. Each EHT pipeline is packaged into an independent container that automates installation, dependency setup, environment configuration, and execution. The containers include our scripts and auxiliary files for conducting the image postprocessing steps, which are unavailable in the original EHT repository.
All four containers are publicly available in a Docker Hub. Additional documentation for deploying and using these containers is available in Github, along with the scripts to generate the figures reproduced in this article. These materials augment existing containers in the EHT Docker Hub and the EHT repositories.

---

**FIGURE 6.** Final images obtained from (a) the original publication from Figure 11 in the EHT Collaboration article and (b) our reproduced results.

---

1. https://hub.docker.com/repository/docker/globalcomputinglab/reproducibility-eht
2. https://github.com/TauferLab/Reproducibility_EHT
3. https://hub.docker.com/u/eventhorizontelescope

---

**REPRODUCIBLE RESEARCH**
We tested the containerized pipelines on commodity hardware (a laptop with an Intel CPU) and a Power9 cluster at the University of Tennessee, Knoxville. Figure 6 compares our results: Figure 6(a) shows the original images from Figure 11 in the EHT Collaboration publication, and Figure 6(b) shows our reproduced images using the containerized pipelines. The two figures show that we can reproduce the M87 images for all three pipelines.

The images in Figure 6 provide us with a qualitative comparison. Both sets of images look visually similar in shape and brightness, and the similarity is consistent across pipelines. To perform a quantitative analysis, we compare the "closure" quantities reported in Table 5 in EHT Collaboration article with those reported by our executions of the three pipelines (Table 2). For each day and each pipeline, we compare both \( \Sigma_{CP} \) and \( \Sigma_{DAFCA} \) quantities computed across the top set of parameters. For brevity, we only report the values with 0% systematic uncertainty. We observe consistency between the two sets of results, with no perfect agreement for the EHT-Imaging and SMILI pipelines. We also find a more significant difference between the original and reproduced values for the DIFMAP pipeline; this is consistent with the discussion of the different time averaging used in DIFMAP.

**LESSONS LEARNED AND GUIDELINES**

We compile lessons learned and guidelines to support the reproducibility of scientific projects based on our experience and observations reproducing the M87 black hole images from the EHT project.

**Data Availability**

The unavailability of the raw data at the time of this project made the direct validation of the pipeline input data unfeasible. As a proxy for data validation, we reproduced Figure 1 in EHT Collaboration publication, as this figure captures properties of the data telescope frequency and coverage. We can reproduce most of these properties, except for the intrasite EHT interferometer baselines (short baselines) because the intra-ALMA data are not available. While this was not the case in this study, any incomplete or missing dataset may result in the user's inability to verify the data integrity thoroughly and can threaten the entire reproducibility process. Data size or ownership constraints can be obstacles to making raw data available to the public. Under these circumstances, data integrity mechanisms such as hashes ensure the correctness of processed data when releasing the raw data is not feasible. We add the additional service to run an MD5 integrity check for the pipeline input data as part of our EHT container set to facilitate data integrity validation.

**Software Availability**

Several pieces of software were unavailable at different stages of the EHT workflow and for the three pipelines. The raw data and corresponding software to process the data were not available at the time of this work. Also, there were no scripts to run the data validation. Therefore, we developed those scripts as well. The code for running the three pipelines is entirely available. Still, the image postprocessing scripts are not, which forced us to experiment with different settings to obtain results comparable to the original for each pipeline. Finally, the plotting libraries used to reproduce the results in Figure 1 from the EHT Collaboration publication were unclear. Thus, we manually tuned our plotting scripts to obtain a suitable plotting configuration. The qualitative differences between the original and reproduced images can result from our manual tuning, indicating that sharing the core for the three pipelines is insufficient to reproduce the original images of the black hole in the galaxy M87. To support portability across platforms, we generated four containers that allow users to execute the data integrity validation and original pipeline codes. We also enable the execution of the end-to-end workflow by providing all auxiliary materials for image postprocessing, figure generation, and result analysis.

**Documentation Availability**

In general, there is insufficient documentation on how to package, install, and execute the EHT pipelines and on performing both qualitative and quantitative analyses of the results. This hinders the overall reproducibility effort. For instance, documentation is key to reproducing Table 5 in the EHT Collaboration publication. Regarding the pipelines, there is insufficient information about software dependencies and versions used as well as file locations and their use. Documenting the whole EHT workflow beyond its image reconstruction components is crucial for the successful reproducibility of the results. Our documentation covering configuration and use of the entire EHT workflow was instrumental to our reproducibility study's success.

**Software Packaging**

The incomplete documentation resulted in installation, dependency, and portability challenges. We manually edited dependencies to allow installation and
compilation and had to override installation instructions resulting in unstable environments. We found that, by containerizing the workflow, we can hide these challenges from the end user, simplifying the installation and deployment of the EHT workflow.

Methods Description
Incomplete descriptions of the results analysis process (e.g., the data averaging time or the additional systematic error budget added to the uncertainties) compromise the reproducibility of the $\chi^2$ statistics in Table 5 in the EHT Collaboration article, as errors add up quickly. Conducting an adequate quantitative assessment of the final results becomes very challenging under these circumstances. In conversations with the authors of this article, members of the EHT Collaboration highlighted how a qualitative comparison of the images is still a good, interpretable solution. Ultimately, different library versions, imaging parameters, or pipelines lead to differences in the reproduced images that are still less than the uncertainty quantified in the EHT Collaboration publication.

Access to Final Results
The EHT Collaboration did not release the output fiducial images. Therefore, we did not have a fixed reference for directly comparing our reproduced and original images. This, in addition to partial access to the data and an incomplete description of the methods used, prevents us from validating the reproduced images thoroughly.

Access to Distributed Knowledge
The EHT Collaboration made substantial investments to allow independent users to qualitatively and quantitatively reproduce their results and ensure the robustness of the EHT project. Nonetheless, we found it challenging to reproduce the original results without direct knowledge of the methods and analyses or the direct collaboration with the studies’ authors. Our experience illustrates the general challenges that users external to a project face when gathering knowledge on data, code, and documentation initially generated from multiple teams in a distributed fashion. The effort of the EHT Collaboration to remove biases by designing and deploying three completely separate pipelines, while instrumental for the trustworthiness of the project results, is also an obstacle to the project’s reproducibility.

CONCLUSION
In this article, we deliver our experience reproducing the black hole images from the EHT project and report new guidance and practices for building reproducible scientific research. Our work complements the work of the EHT Collaboration with supplemental data, scripts, documentation, and a set of containers. Postdocs, graduate and undergraduate students, and even high school students can benefit from accessing our data and code and using our documentation to reproduce findings from the EHT project; learn about the EHT funding; and ultimately get involved in science, technology, engineering, and mathematics (STEM) research. Our guidance and practices can be incorporated more broadly by other scientific workflows. The EHT project continues to be a leader in reproducibility efforts and has provided comprehensive data products for their recent observations of Sagittarius A*.

Assessing the level of detail required to cover the vast knowledge developed in a project the size of EHT

---

**TABLE 2.** Closure quantity $\chi^2$ values and statistics for top set images with 0% systematic uncertainty.

|                | DIFMAP |               | EHT-Imaging |               | SMILI |               |
|----------------|--------|---------------|-------------|---------------|-------|---------------|
|                | Original | Reproduced | Original | Reproduced | Original | Reproduced |
| 5 April top set | $\chi^2_{CTP}$ | 9.40 ± 3.35 | 6.81 (δ = 2.59) | 1.00 ± 0.13 | 1.03 (δ = 0.03) | 1.09 ± 0.21 | 1.08 (δ = 0.01) |
|                | $\chi^2_{logCA}$ | 4.99 ± 1.17 | 35.12 (δ = 30.13) | 0.97 ± 0.27 | 1.38 (δ = 0.41) | 1.11 ± 0.28 | 1.29 (δ = 0.18) |
| 6 April top set | $\chi^2_{CTP}$ | 4.40 ± 1.45 | 2.84 (δ = 1.56) | 1.59 ± 0.16 | 0.88 (δ = 0.71) | 1.49 ± 0.23 | 1.30 (δ = 0.19) |
|                | $\chi^2_{logCA}$ | 3.49 ± 1.51 | 24.43 (δ = 20.94) | 1.00 ± 0.14 | 1.21 (δ = 0.21) | 1.22 ± 0.25 | 1.62 (δ = 0.40) |
| 10 April top set | $\chi^2_{CTP}$ | 3.93 ± 2.10 | 2.36 (δ = 1.57) | 0.83 ± 0.10 | 0.99 (δ = 0.16) | 0.75 ± 0.14 | 1.24 (δ = 0.49) |
|                | $\chi^2_{logCA}$ | 1.57 ± 0.64 | 22.74 (δ = 21.17) | 1.30 ± 0.17 | 0.82 (δ = 0.48) | 1.11 ± 0.32 | 0.86 (δ = 0.25) |
| 11 April top set | $\chi^2_{CTP}$ | 4.01 ± 1.67 | 1.91 (δ = 2.1) | 0.97 ± 0.10 | 0.88 (δ = 0.09) | 1.23 ± 0.28 | 1.00 (δ = 0.23) |
|                | $\chi^2_{logCA}$ | 3.33 ± 1.01 | 61.32 (δ = 57.99) | 0.99 ± 0.13 | 0.85 (δ = 0.14) | 1.14 ± 0.13 | 1.04 (δ = 0.1) |

We compute the difference δ between the top set values in Table 5 in the EHT Collaboration article and our reproduced values. None of the values agrees exactly, but our values are consistent with the spread reported by the EHT Collaboration for the EHT-Imaging and SMILI pipelines.
is a complex task. Finding the balance between the effort from original research teams to enable reproducibility and users attempting to reproduce the results is still an open question. Our experience with the EHT and LIGO projects reveals an essential and recurring issue in reproducibility: challenges remain in disseminating findings in a way that allows the reproducibility of results without direct interaction with the original team that produced them.

Understanding how reproducibility is incorporated in astrophysics workflows and sharing practices in reproducible scientific software can enable more robust science across disciplines and help accelerate the pace of scientific discovery. Enabling robust computational research allows scientists to reproduce published findings, making their methods visible and accessible to a much larger audience. One can imagine researchers and students at various levels of education accessing the code, data, and workflow information and being able to regenerate findings, learn about scientific methods, and engage in STEM research. This is already happening in gravitational wave physics, and such practices can be incorporated more broadly across science domains.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant 2041977, Grant 1941443, Grant 2041901, Grant 1664162, and Grant 2041878. We thank Kazunori Akiyama, Lindy Blackburn, Chi-kwan Chan, and Rebecca White for their helpful discussions. The authors would like to thank the reviewers for their feedback, in particular Oliver Porth (https://doi.org/10.22541/au.16661871.15772020/v1) and EHT Collaboration members Katherine Bouman, Jose L. Gómez, Kotaro Moriyama, and Kazunori Akiyama (https://doi.org/10.22541/au.1669771774.46972022/v1).

REFERENCES

1. The Event Horizon Telescope Collaboration et al., “First M87 Event Horizon Telescope results. VI. The shadow and mass of the central black hole,” Astrophys. J. Lett., vol. 875, no. 1, p. L6, Apr. 2019, doi: 10.3847/2041-8213/ab1141.
2. D. A. Brown, K. Vahi, M. Tauber, V. Welch, and E. Deelman, “Reproducing GW150914: The first observation of gravitational waves from a binary black hole merger,” Comput. Sci. Eng., vol. 23, no. 2, pp. 73–82, Mar./Apr. 2021, doi: 10.1109/MCSE.2021.3059232.
3. The Event Horizon Telescope Collaboration et al., “First M87 Event Horizon Telescope results. IV. Imaging the central supermassive black hole,” Astrophys. J. Lett., vol. 875, no. 1, p. L4, Apr. 2019, doi: 10.3847/2041-8213/ab0e88.
4. The Event Horizon Telescope Collaboration et al., “First M87 Event Horizon Telescope results. III. Data processing and calibration,” Astrophys. J. Lett., vol. 875, no. 1, p. L3, Apr. 2019.
5. The Event Horizon Telescope Collaboration et al., “First M87 EHT results: Calibrated data,” CyVerse Data Commons, 2019. Accessed: Apr. 30, 2022. [Online]. Available: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/EHTC_FirstM87Results_Apr2019.
6. The Event Horizon Telescope Collaboration et al., “First M87 EHT results: Imaging pipelines.” GitHub. Accessed: Apr. 30, 2022. [Online]. Available: https://github.com/eventhorizontelescope/2019-D01-02
7. P. Arras, P. Frank, and P. Haim, “Variable structures in M87* from space, time and frequency resolved interferometry,” Nature Astron., vol. 6, no. 2, pp. 259–269, Feb. 2022, doi: 10.1038/s41550-021-01548-0.
8. C. L. Carilli and N. Thayagarajan, “Hybrid mapping of the black hole shadow in M87,” Astrophys. J., vol. 924, no. 2, p. 125, Jan. 2022, doi: 10.3847/1538-4357/ac3cb.
9. W. Lockhart and S. E. Gralla, “How narrow is the M87* ring? I. The choice of closure likelihood function,” Monthly Notices Roy. Astron. Soc., vol. 509, no. 3, pp. 3643–3659, Nov. 2021, doi: 10.1093/mnras/stab3204.
10. M. Miyoshi, Y. Kato, and J. Makino, “Comments on ’Imaging reanalyses of EHT data’,” 2022. [Online]. Available: https://arxiv.org/pdf/2207.13279.pdf
11. R. Kieron et al., “A case study in scientific reproducibility from the Event Horizon Telescope (EHT),” in Proc. IEEE 17th Int. Conf. eScience (eScience), 2021, pp. 249–250, doi: 10.1109/eScience51609.2021.00045.
12. The Event Horizon Telescope Collaboration et al., “First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole,” Astrophys. J. Lett., vol. 875, no. 1, p. L1, Apr. 2019, doi: 10.3847/2041-8213/ab0ec7.
13. L. Blackburn et al., “EHT-HOPS pipeline for millimeter VLBI data reduction,” Astrophys. J., vol. 882, no. 1, p. 23, Aug. 2019, doi: 10.3847/1538-4357/ab328d.
14. The Event Horizon Telescope Collaboration, “Public data release of the Event Horizon Telescope 2017 observations,” CyVerse Data Commons, 2022. Accessed: Dec. 30, 2022. [Online]. Available: https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/EHTC_2017L1_May2022.
RIA PATEL is a fourth-year undergraduate student majoring in computer science and minoring in mathematics and physics at the University of Tennessee, Knoxville, Knoxville, TN, USA. She is an undergraduate research assistant in the Global Computing Lab under Dr. Taufer. She is a Member of IEEE. Contact her at rpatel77@vols.utk.edu.

BRANDAN ROACHELL is an undergraduate student majoring in computer science and minoring in mathematics at the University of Tennessee, Knoxville, Knoxville, TN, USA. He is an undergraduate research assistant in the Global Computing Lab under Dr. Taufer. Contact him at broachel@vols.utk.edu.

SILVINA CAINO-LORES is a research assistant professor at the University of Tennessee–Knoxville, Knoxville, TN, USA. Her research interests include cloud computing, in-memory computing and storage, high-performance computing scientific simulations, and datacentric paradigms. Caino-Lores received her Ph.D. degree in computer science and technology from Carlos III University of Madrid. She is a Member of IEEE. Contact her at scaninolo@utk.edu.

ROSS KETRON is an undergraduate student in computer science at the University of Tennessee, Knoxville, Knoxville, TN, USA, and an undergraduate research assistant in the Global Computing Lab under Dr. Taufer. Contact him at cketron2@vols.utk.edu.

JACOB LEONARD is an undergraduate student in computer science at the University of Tennessee, Knoxville, Knoxville, TN, USA, and an undergraduate research assistant in the Global Computing Lab under Dr. Taufer. He is a Member of IEEE. Contact him at jleona19@vols.utk.edu.

NIGEL TAN is a graduate student in computer science at the University of Tennessee, Knoxville, Knoxville, TN, USA. His research interests include software portability across heterogeneous platforms. Contact him at ntan1@vols.utk.edu.

KARAN VAHI is a senior computer scientist at the University of Southern California (USC) Information Sciences Institute, Los Angeles, CA, USA. His research interests include scientific workflows and distributed computing systems. Vahi received his M.S. degree in computer science in 2003 from USC. He is a Member of IEEE. Contact him at vahi@isi.edu.

DUNCAN A. BROWN is the Charles Brightman Professor of Physics at Syracuse University, Syracuse, NY, USA. His research interests include gravitational wave astronomy and astrophysics and large-scale scientific workflows. Brown received his Ph.D. degree in physics from the University of Wisconsin-Milwaukee. He was a member of the LIGO Scientific Collaboration from 1999 to 2018 and is a fellow of the American Physical Society. Contact him at dabrown@syr.edu.

EWA DEELMAN is a research director at the University of Southern California (USC) Information Sciences Institute, Los Angeles, CA, USA; is a research professor at the USC Computer Science Department; and leads the design and development of the Pegasus Workflow Management software. Her research interests include the interplay between automation and the management of scientific workflows that include resource provisioning and data management. Deelman received her Ph.D. degree in computer science from Rensselaer Polytechnic Institute. Contact her at deelman@isi.edu.

MICHELA TAUFER holds the Jack Dongarra Professorship in High-Performance Computing within the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville, Knoxville, TN, USA. Her research interests include the intersection of computational sciences, high-performance computing, and data analytics. Tauber received her Ph.D. degree in computer science from the Swiss Federal Institute of Technology (EHT) in 2002. Contact her at taufer@utk.edu.