Data Analysis Challenges for Multi-Messenger Astrophysics

Peter S. Shawhan,[1,2] Patrick R. Brady,[3] Adam Brazier,[4] S. Bradley Cenko,[5,2] Mario Jurić,[6] and Erik Katsavounidis[7]

[1]*Department of Physics, University of Maryland, College Park, MD, USA;* pshawhan@umd.edu
[2]*Joint Space-Science Institute, Maryland, USA*
[3]*Center for Gravitation, Cosmology and Astrophysics, University of Wisconsin–Milwaukee, WI, USA*
[4]*Department of Astronomy, Cornell University, Ithaca, NY, USA*
[5]*NASA Goddard Space Flight Center, Greenbelt, MD, USA*
[6]*Department of Astronomy, University of Washington, Seattle, WA, USA*
[7]*LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA*

Abstract. Recent multi-messenger observations of gravitational-wave and high-energy neutrino sources together with electromagnetic signatures have opened new ways of observing the Universe. These promise a future in which physics and astronomy will be advanced by combining observations and data from across the electromagnetic spectrum with gravitational waves and neutrinos. We consider the challenges the field is facing in fully utilizing data for multi-messenger astrophysics. Such data come from heterogeneous detector networks and standards, and their analysis is often time-critical to guide further observations. In this area, science capabilities depend on the interplay among observation, theory and computational/modeling work. Advances in data science and computing present additional opportunities and considerations in analyzing such data. We invited ADASS participants to a Birds of a Feather session to engage in discussion on the challenges and opportunities in data analysis for multi-messenger astrophysics.

1. Introduction

Until recently, mankind relied almost exclusively on electromagnetic (EM) waves (or photons) spanning the spectrum to reveal the content and workings of the universe. Large, highly sensitive neutrino and gravitational-wave observatories have now made additional astrophysical "messengers" available for study, with initial discoveries (Aartsen et al. 2013; Abbott et al. 2016) progressing to understanding source populations (Aartsen et al. 2018; LIGO Scientific Collaboration & Virgo Collaboration 2018b,a).

Furthermore, highly energetic sources such as neutron star binary mergers, stellar core collapse, and relativistic jets can produce multiple types of emissions. The relatively new field of *multi-messenger astrophysics* involves combining observations of the same events with different messengers, taking advantage of their different characteristics to collect and relate information about the core "engine", outflows and environment of individual events or of populations. Modern facilities and instruments have strengths to reveal different properties. For instance, gamma-ray and neutrino fluxes and spectra can reveal particle acceleration; X-ray and radio afterglows can localize sources and characterize their environments; gravitational waves can measure binary mass and orientation parameters and distance; and visible/infrared can precisely localize events, provide redshifts, and trace out expansion and thermal signatures.

The idea for this Birds of a Feather (BoF) session came out of a workshop held at the University of Maryland in May 2018. That Workshop on Cyberinfrastructure for Multi-Messenger Astrophysics (CiMMA 2018) brought together about 20 experts in astrophysics, computational science, and everywhere in between to discuss how cyberinfrastructure—broadly interpreted—can and should be used more effectively to enable multi-messenger science. The opportunities and challenges considered during the workshop are discussed in the workshop report (Allen et al. 2018).

At around the same time, the U.S. National Science Foundation's Office of Advanced Cyberinfrastructure encouraged proposals to its Cyberinfrastructure for Emerging Science and Engineering Research (CESER) program for "scalable data-driven cyberinfrastructure exemplars that will accelerate discovery for one or more science and engineering research communities". We and our colleagues recognized the ideal match to the goals and needs of multi-messenger astrophysics—bringing together two of NSF's Big Ideas—and launched the SCiMMA Project: Community planning for Scalable Cyberinfrastructure to support Multi-Messenger Astrophysics.[1] The ADASS BoF session provided an ideal venue to share this vision and gain valuable input from experts in attendance.

2. Science Drivers

The 2017 gravitational-wave detection of the binary neutron star merger GW170817 (Abbott et al. 2017a), accompanied by a gamma-ray burst and followed up thoroughly using EM facilities around the world (Abbott et al. 2017b), was an outstanding inauguration of multi-messenger astrophysics which yielded many important insights. However, it is just one example. The following scenarios, which could plausibly arise over the next several years, give a sense of other science opportunities:

- Early warning of a nearby compact binary merger, via gravitational waves, allows the earliest phase of the EM counterpart to be identified and measured.
- High-energy neutrinos detected and localized to a galaxy cluster trigger EM follow-up and the observation of a tidal disruption event.
- Gravitational waves from supermassive black-hole binaries detected by pulsar timing arrays allow studies of galaxy properties and accretion disk physics.
- A Galactic or Local Group supernova is observed in all the messengers!

Initial detection of these events will likely be done by "anchor" facilities such as LIGO, IceCube, wide-field gamma-ray survey missions, LSST, Super/Hyper-Kamiokande, and pulsar timing arrays, but follow-up observations by many other facilities will be crucial to fully understand these systems. Efficient, robust searches for signals must be supplemented with rapid characterization, photometric analysis, signal modeling and multi-dimensional parameter estimation to extract astrophysics from the combined observational data.

[1] `https://scimma.org/`

3. Cyberinfrastructure for Multi-Messenger Astrophysics

In this context we consider *cyberinfrastructure* to include distributed data-handling, computing, analysis, and collaboration services/systems to enable discovery, education, and innovation. Of course, many advanced computing systems and techniques are in place to serve individual projects, and various communication and information-sharing tools already exist, such as SNEWS, AMON, GCN, TNS, SNEx, ATel, ANTARES, and VOEvent. Survey data and catalogs of known objects are also important for identifying and classifying transients.

Multi-messenger astronomy, by definition, involves diverse facilities. Doing it well presents distinct challenges due to having highly heterogeneous facilities, data, and people; high-volume, high-velocity transient streams; rapidly developing, dynamic collaborations; heterogeneous data sharing policies; competition for follow-up resources; the need for rapid modeling and intelligent scheduling; and tension between human versus machine communications. To realize the science potential of multi-messenger astrophysics, we expect the community to need capabilities such as:

- a scalable framework to facilitate joint analysis, enabling teams to work together, while respecting different scientific cultures
- real-time decision making in event observation and follow-up
- coordination of observing resources, with capability-based access controls
- sustainable, long-term archival storage with standardization of data sets
- data escrow, pre-registration of analyses
- machine-readable communication standards and protocols

Pursuing this also offers many interesting avenues for research and development of computer/data science, including machine learning (deep learning), purpose-built hardware for real-time inference, data access/analysis with differential privacy, uncertainty quantification and predictive modeling, handling of missing or imbalanced data, and resource optimization.

The general goal of the SCiMMA Project is to identify the key questions and cyberinfrastructure projects required by the community to take full advantage of current facilities and imminent next-generation projects for multi-messenger astrophysics. Using input from the general community and specific stakeholder groups, the project aims to produce a community white paper documenting needs and opportunities, and a strategic plan for an institute to address these needs, by the middle of 2019.

Join the conversation! Visit the scimma.org web site, join the "scimma" Google Groups forum, attend upcoming workshops, and contribute to activity areas of interest to you (data management, application integration, machine learning, planning observations, astrophysical inference, modeling and theory, education and workforce development, etc.) We are endeavoring to understand how this kind of distributed work can be coordinated effectively through an institute to support excellent science in the future.

4. Discussion

Audience questions and discussion in the BoF session provided useful input. Notable lines of discussion included:

- The role of inter-agency task forces or other coordination, within the U.S. and internationally, for setting consistent priorities and complementary funding.

• We should absorb and build on, as much as possible, the activities and lessons learned from the ASTERICS project,[2] which has been underway in Europe since 2015 and is completing in 2019.

• Consider what systems already exist for communication and coordiation of robotic telescopes, worldwide. How can we build on it?

• Some multi-messenger observations are open while others will involve proprietary data. Part of this effort is to understand the policy issues and to enable collaborations to form quickly, avoiding a slow ad-hoc memorandum-of-understanding process while still respecting data access restrictions.

• Besides detecting potentially interesting events, it is important to quickly evaluate them and make judgments about using additional observing resources.

• How do we get relevant people involved, and how can we ensure that useful tools are actually used? Need to engage people and existing systems, and consider the sociological and educational aspects as much as the technical ones.

• How can we work effectively with aggregate data sets which come from different instruments? Do we attempt to host data in one place, or do distributed processing? Can we find and refer to all the relevant data with one handle?

• The ability to filter information streams is important, using "brokers" or other mechanisms.

We look forward to future input and discussion on these and other topics.

Acknowledgments. We acknowlege the participants at the CiMMA 2018 workshop and all contributors to the ongoing SCiMMA Project for useful discussions, and we thank everyone who attended the ADASS Birds of a Feather session for their input. This planning and community engagement effort is supported by the U.S. National Science Foundation through grants PHY-1838082 and OAC-1841625.

References

Aartsen, M., et al. 2018, Advances in Space Research, 62, 2902 . Origins of Cosmic Rays, URL http://www.sciencedirect.com/science/article/pii/S0273117717303757
Aartsen, M. G., et al. 2013, Phys.Rev.Lett, 111, 021103. 1304.5356
Abbott, B. P., et al. 2016, Phys.Rev.Lett, 116, 061102. 1602.03837
— 2017a, Phys.Rev.Lett, 119, 161101. 1710.05832
— 2017b, ApJ, 848, L12. 1710.05833
Allen, G., Anderson, W., Blaufuss, E., Bloom, J. S., Brady, P., Burke-Spolaor, S., Cenko, S. B., Connolly, A., Couvares, P., Fox, D., Gal-Yam, A., Gezari, S., Goodman, A., Grant, D., Groot, P., Guillochon, J., Hanna, C., Hogg, D. W., Holley-Bockelmann, K., Howell, D. A., Kaplan, D., Katsavounidis, E., Kowalski, M., Lehner, L., Muthukrishna, D., Narayan, G., Peek, J. E. G., Saha, A., Shawhan, P., & Taboada, I. 2018, ArXiv e-prints, arXiv:1807.04780. 1807.04780
LIGO Scientific Collaboration, & Virgo Collaboration 2018a, arXiv e-prints, arXiv:1811.12940. 1811.12940
— 2018b, arXiv e-prints, arXiv:1811.12907. 1811.12907

[2]https://www.asterics2020.eu/