The Astrophysical Journal, 897:139 (38pp), 2020 July 10 https://doi.org/10.3847/1538-4357/ab91a4

THEMIS: A Parameter Estimation Framework for the Event Horizon Telescope

Avery E. Broderick[1,2,3], Roman Gold[1,4], Mansour Karami[1,2], Jorge A. Preciado-López[1], Paul Tiede[1,2,3], Hung-Yi Pu[1], Kazunori Akiyama[5,6,7,8], Antxon Alberdi[9], Walter Alef[10], Keiichi Asada[11], Rebecca Azulay[10,12,13], Anne-Kathrin Baczko[10], Mislav Baloković[8,14], John Barrett[6], Dan Bintley[15], Lindy Blackburn[8,14], Wilfred Boland[16], Katherine L. Bouman[8,14,17], Geoffrey C. Bower[18], Michael Bremer[19], Christiaan D. Brinkerink[20], Roger Brissenden[8,14], Silke Britzen[10], Dominique Broguiere[19], Thomas Bronzwaer[20], Do-Young Byun[21,22], John E. Carlstrom[23,24,25,26], Andrew Chael[8,14], Shami Chatterjee[27], Koushik Chatterjee[28], Ming-Tang Chen[18], Yongjun Chen (陈永军)[29,30], Ilje Cho[21,22], John E. Conway[31], James M. Cordes[27], Geoffrey B. Crew[6], Yuzhu Cui[32,33], Jordy Davelaar[20], Mariafelicia De Laurentis[4,34,35], Roger Deane[36,37], Jessica Dempsey[15], Gregory Desvignes[38,10], Sheperd S. Doeleman[8,14], Ralph P. Eatough[10], Heino Falcke[20], Vincent L. Fish[6], Ed Fomalont[5], Raquel Fraga-Encinas[20], Per Friberg[15], Christian M. Fromm[4], Peter Galison[8,39,40], Charles F. Gammie[41,42], Roberto García[19], Olivier Gentaz[19], Boris Georgiev[1,2,3], Ciriaco Goddi[20,43], José L. Gómez[9], Minfeng Gu (顾敏峰)[29,44], Mark Gurwell[14], Kazuhiro Hada[32,33], Michael H. Hecht[6], Ronald Hesper[45], Luis C. Ho (何子山)[46,47], Paul Ho[11], Mareki Honma[32,33], Chih-Wei L. Huang[11], Lei Huang (黄磊)[29,44], David H. Hughes[48], Makoto Inoue[11], Sara Issaoun[20], David J. James[8,14], Michael Janssen[20], Britton Jeter[1,2,3], Wu Jiang (江悟)[29], Alejandra Jiménez-Rosales[49], Michael D. Johnson[8,14], Svetlana Jorstad[64,51], Taehyun Jung[21,22], Ramesh Karuppusamy[10], Tomohisa Kawashima[7], Garrett K. Keating[14], Mark Kettenis[52], Jae-Young Kim[10], Jongsoo Kim[21], Motoki Kino[7,53], Jun Yi Koay[11], Patrick M. Koch[11], Shoko Koyama[11], Michael Kramer[10], Carsten Kramer[19], Thomas P. Krichbaum[10], Cheng-Yu Kuo[54], Sang-Sung Lee[21], Yan-Rong Li (李彦荣)[55], Zhiyuan Li (李志远)[56,57], Michael Lindqvist[31], Rocco Lico[10], Kuo Liu[10], Elisabetta Liuzzo[58], Wen-Ping Lo[11,59], Andrei P. Lobanov[10], Laurent Loinard[60,61], Colin Lonsdale[6], Ru-Sen Lu (路如森)[10,29], Nicholas R. MacDonald[10], Jirong Mao (毛基荣)[62,50,63], Alan P. Marscher[64], Iván Martí-Vidal[12,13], Satoki Matsushita[11], Lynn D. Matthews[6], Karl M. Menten[10], Yosuke Mizuno[4], Izumi Mizuno[15], James M. Moran[8,14], Kotaro Moriyama[6,32], Monika Moscibrodzka[20], Cornelia Müller[10,20], Hiroshi Nagai[7,33], Neil M. Nagar[65], Masanori Nakamura[11], Ramesh Narayan[8,14], Gopal Narayanan[66], Iniyan Natarajan[37], Roberto Neri[19], Chunchong Ni[1,2,3], Aristeidis Noutsos[10], Hiroki Okino[32,67], Héctor Olivares[4], Gisela N. Ortiz-León[10], Tomoaki Oyama[32], Daniel C. M. Palumbo[8,14], Jongho Park[11], Ue-Li Pen[1,68,69,70], Dominic W. Pesce[8,14], Vincent Piétu[19], Richard Plambeck[71], Aleksandar PopStefanija[66], Oliver Porth[4,28], Ben Prather[41], Venkatessh Ramakrishnan[65], Ramprasad Rao[18], Mark G. Rawlings[15], Alexander W. Raymond[8,14], Luciano Rezzolla[4], Bart Ripperda[4,72,73], Freek Roelofs[20], Alan Rogers[6], Eduardo Ros[10], Mel Rose[74], Helge Rottmann[10], Chet Ruszczyk[6], Benjamin R. Ryan[75,76], Kazi L. J. Rygl[58], Salvador Sánchez[77], David Sánchez-Arguelles[48,78], Mahito Sasada[32,79], Tuomas Savolainen[10,80,81], F. Peter Schloerb[66], Karl-Friedrich Schuster[19], Lijing Shao[10,47], Zhiqiang Shen (沈志强)[29,30], Des Small[52], Bong Won Sohn[21,22,82], Jason SooHoo[6], Fumie Tazaki[32], Remo P. J. Tilanus[20,43,83], Michael Titus[6], Kenji Toma[84,85], Pablo Torne[10,77], Efthalia Traianou[10], Sascha Trippe[86], Shuichiro Tsuda[32], Ilse van Bemmel[52], Huib Jan van Langevelde[52,87], Daniel R. van Rossum[20], Jan Wagner[10], John Wardle[88], Jonathan Weintroub[8,14], Norbert Wex[10], Robert Wharton[10], Maciek Wielgus[8,14], George N. Wong[41], Qingwen Wu (吴庆文)[89], Doosoo Yoon[28], André Young[20], Ken Young[14], Ziri Younsi[4,90], Feng Yuan (袁峰)[29,44,91], Ye-Fei Yuan (袁业飞)[92], J. Anton Zensus[10], Guangyao Zhao[9,21], Shan-Shan Zhao[20,56], and Ziyan Zhu[40]

(The Event Horizon Telescope Collaboration)

[1] Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada; abroderick@perimeterinstitute.ca
[2] Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada
[3] Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[4] Institut für Theoretische Physik, Goethe-Universität Frankfurt, Max-von-Laue-Straße 1, D-60438 Frankfurt am Main, Germany
[5] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[6] Massachusetts Institute of Technology Haystack Observatory, 99 Millstone Road, Westford, MA 01886, USA
[7] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[8] Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA
[9] Instituto de Astrofísica de Andalucía-CSIC, Glorieta de la Astronomía s/n, E-18008 Granada, Spain
[10] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany
[11] Institute of Astronomy and Astrophysics, Academia Sinica, 11F of Astronomy-Mathematics Building, AS/NTU No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, R.O.C
[12] Departament d'Astronomia i Astrofísica, Universitat de València, C. Dr. Moliner 50, E-46100 Burjassot, València, Spain
[13] Observatori Astronòmic, Universitat de València, C. Catedrático José Beltrán 2, E-46980 Paterna, València, Spain
[14] Center for Astrophysics, Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[15] East Asian Observatory, 660 N. A'ohoku Place, Hilo, HI 96720, USA
[16] Nederlandse Onderzoekschool voor Astronomie (NOVA), P.O. Box 9513, 2300 RA Leiden, The Netherlands
[17] California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA
[18] Institute of Astronomy and Astrophysics, Academia Sinica, 645 N. A'ohoku Place, Hilo, HI 96720, USA

[19] Institut de Radioastronomie Millimétrique, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France
[20] Department of Astrophysics, Institute for Mathematics, Astrophysics and Particle Physics (IMAPP), Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[21] Korea Astronomy and Space Science Institute, Daedeok-daero 776, Yuseong-gu, Daejeon 34055, Republic of Korea
[22] University of Science and Technology, Gajeong-ro 217, Yuseong-gu, Daejeon 34113, Republic of Korea
[23] Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[24] Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[25] Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[26] Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[27] Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA
[28] Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
[29] Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, People's Republic of China
[30] Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China
[31] Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, SE-43992 Onsala, Sweden
[32] Mizusawa VLBI Observatory, National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu, Iwate 023-0861, Japan
[33] Department of Astronomical Science, The Graduate University for Advanced Studies (SOKENDAI), 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[34] Dipartimento di Fisica "E. Pancini," Universitá di Napoli "Federico II," Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
[35] INFN Sez. di Napoli, Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
[36] Department of Physics, University of Pretoria, Lynnwood Road, Hatfield, Pretoria 0083, South Africa
[37] Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa
[38] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France
[39] Department of History of Science, Harvard University, Cambridge, MA 02138, USA
[40] Department of Physics, Harvard University, Cambridge, MA 02138, USA
[41] Department of Physics, University of Illinois, 1110 West Green Street, Urbana, IL 61801, USA
[42] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA
[43] Leiden Observatory—Allegro, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[44] Key Laboratory for Research in Galaxies and Cosmology, Chinese Academy of Sciences, Shanghai 200030, People's Republic of China
[45] NOVA Sub-mm Instrumentation Group, Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
[46] Department of Astronomy, School of Physics, Peking University, Beijing 100871, People's Republic of China
[47] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
[48] Instituto Nacional de Astrofísica, Óptica y Electrónica. Apartado Postal 51 y 216, 72000. Puebla Pue., México
[49] Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstr. 1, D-85748 Garching, Germany
[50] Center for Astronomical Mega-Science, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing, 100012, People's Republic of China
[51] Astronomical Institute, St. Petersburg University, Universitetskij pr., 28, Petrodvorets,198504 St.Petersburg, Russia
[52] Joint Institute for VLBI ERIC (JIVE), Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands
[53] Kogakuin University of Technology & Engineering, Academic Support Center, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan
[54] Physics Department, National Sun Yat-Sen University, No. 70, Lien-Hai Road, Kaosiung City 80424, Taiwan, R.O.C
[55] Key Laboratory for Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing, People's Republic of China
[56] School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China
[57] Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, Nanjing 210023, People's Republic of China
[58] Italian ALMA Regional Centre, INAF-Istituto di Radioastronomia, Via P. Gobetti 101, I-40129 Bologna, Italy
[59] Department of Physics, National Taiwan University, No.1, Sect. 4, Roosevelt Road, Taipei 10617, Taiwan, R.O.C
[60] Instituto de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, Morelia 58089, México
[61] Instituto de Astronomía, Universidad Nacional Autónoma de México, CdMx 04510, México
[62] Yunnan Observatories, Chinese Academy of Sciences, 650011 Kunming, Yunnan Province, People's Republic of China
[63] Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, 650011 Kunming, People's Republic of China
[64] Institute for Astrophysical Research, Boston University, 725 Commonwealth Avenue, Boston, MA 02215, USA
[65] Astronomy Department, Universidad de Concepción, Casilla 160-C, Concepción, Chile
[66] Department of Astronomy, University of Massachusetts, 01003, Amherst, MA, USA
[67] Department of Astronomy, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[68] Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[69] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[70] Canadian Institute for Advanced Research, 180 Dundas Street West, Toronto, ON M5G 1Z8, Canada
[71] Radio Astronomy Laboratory, University of California, Berkeley, CA 94720, USA
[72] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
[73] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[74] Steward Observatory and Department of Astronomy, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
[75] CCS-2, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545, USA
[76] Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
[77] Instituto de Radioastronomía Milimétrica, IRAM, Avenida Divina Pastora 7, Local 20, E-18012, Granada, Spain
[78] Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, 03940, Ciudad de México, México
[79] Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
[80] Aalto University, Department of Electronics and Nanoengineering, PL 15500, FI-00076 Aalto, Finland
[81] Aalto University Metsähovi Radio Observatory, Metsähovintie 114, FI-02540 Kylmälä, Finland
[82] Department of Astronomy, Yonsei University, Yonsei-ro 50, Seodaemun-gu, 03722 Seoul, Republic of Korea
[83] Netherlands Organisation for Scientific Research (NWO), Postbus 93138, 2509 AC Den Haag, The Netherlands
[84] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan
[85] Astronomical Institute, Tohoku University, Sendai 980-8578, Japan
[86] Department of Physics and Astronomy, Seoul National University, Gwanak-gu, Seoul 08826, Republic of Korea
[87] Leiden Observatory, Leiden University, Postbus 2300, 9513 RA Leiden, The Netherlands
[88] Physics Department, Brandeis University, 415 South Street, Waltham, MA 02453, USA
[89] School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, People's Republic of China
[90] Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK
[91] School of Astronomy and Space Sciences, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, People's Republic of China

[92] Astronomy Department, University of Science and Technology of China, Hefei 230026, People's Republic of China
Received 2019 September 12; revised 2020 April 29; accepted 2020 May 7; published 2020 July 10

Abstract

The Event Horizon Telescope (EHT) provides the unprecedented ability to directly resolve the structure and dynamics of black hole emission regions on scales smaller than their horizons. This has the potential to critically probe the mechanisms by which black holes accrete and launch outflows, and the structure of supermassive black hole spacetimes. However, accessing this information is a formidable analysis challenge for two reasons. First, the EHT natively produces a variety of data types that encode information about the image structure in nontrivial ways; these are subject to a variety of systematic effects associated with very long baseline interferometry and are supplemented by a wide variety of auxiliary data on the primary EHT targets from decades of other observations. Second, models of the emission regions and their interaction with the black hole are complex, highly uncertain, and computationally expensive to construct. As a result, the scientific utilization of EHT observations requires a flexible, extensible, and powerful analysis framework. We present such a framework, THEMIS, which defines a set of interfaces between models, data, and sampling algorithms that facilitates future development. We describe the design and currently existing components of THEMIS, how THEMIS has been validated thus far, and present additional analyses made possible by THEMIS that illustrate its capabilities. Importantly, we demonstrate that THEMIS is able to reproduce prior EHT analyses, extend these, and do so in a computationally efficient manner that can efficiently exploit modern high-performance computing facilities. THEMIS has already been used extensively in the scientific analysis and interpretation of the first EHT observations of M87.

Unified Astronomy Thesaurus concepts: Astrophysical black holes (98); Galactic center (565); Astronomy data analysis (1858); Very long baseline interferometry (1769); Submillimeter astronomy (1647)

1. Introduction

The Event Horizon Telescope (EHT), a global array of millimeter and submillimeter radio telescopes, has resolved the horizons of at least two black holes (Doeleman et al. 2008, 2009, 2012; Doeleman 2010; Event Horizon Telescope Collaboration et al. 2019a). This provides a unique window on the high-energy astrophysical processes responsible for the substantial growth and inordinate influence of supermassive black holes (Fabian 2012; Heckman & Best 2014), the dynamics and thermodynamics of material in the strong gravity regime (Narayan et al. 1998; Yuan & Narayan 2014), and the fundamental nature of black holes (Broderick et al. 2014; Psaltis et al. 2016). However, efficiently and accurately extracting this information from the observational data presents numerous challenges, requiring the development of novel analysis tools tailored to the EHT data products, EHT-target properties, and auxiliary information.

The EHT achieves an extraordinary resolution of 13 μas, making it the highest resolution imaging instrument in the history of astronomy. It does this via very long baseline interferometry (VLBI), in which information from pairs of individual stations separated by Earth-size distances are combined to measure small-scale structure on the sky. The resulting data take the form of complex visibilities, directly related to the Fourier transform of the image. This can be performed in all four Stokes parameters, yielding complete information about the resolved polarization structures (e.g., Johnson et al. 2015). In the near future, this will be extended to multiple wavelengths (1.3 and 0.87 mm; Falcke 2017). Millimeter-VLBI observations of the primary EHT targets have already been carried out at multiple epochs, covering times ranging from 10 s to 10 yr (Doeleman et al. 2008; Fish et al. 2011, 2016; Johnson et al. 2015).

The first horizon-resolving images have been published, with ancillary scientific analyses (Event Horizon Telescope Collaboration et al. 2019a, 2019b, 2019c, 2019d, 2019e, 2019f, hereafter Papers I, II, III, IV, V, and VI, respectively). These have revealed a resolved shadow with brightness maximum offset from the direction of the large-scale jet, broadly consistent with the model predictions in Broderick & Loeb (2009a), Dexter et al. (2012), and Mościbrodzka et al. (2016). Extracting quantitative information about these has required detailed model comparisons directly with the complex visibilities (see, e.g., Papers V, VI).

Difficulties in the phase calibration, and lesser—though still significant—complications in the amplitude calibration of these visibilities, have motivated the construction of a set of VLBI observables (e.g., visibility amplitudes and closure phases: Jennison 1958; closure amplitudes: Twiss et al. 1960, 1962; visibility polarization fractions: Wardle 1971, Roberts et al. 1994, Fish et al. 2009. Johnson et al. 2015, etc.) that probe the underlying image structure in nonintuitive ways. These have traditionally been interpreted within the context of a simple set of phenomenological models, e.g., multicomponent Gaussians. However, the substantial structure anticipated on horizon scales exhibited by the primary EHT targets has given rise to a broader modeling effort, which includes a variety of physical processes (Narayan et al. 1998; Falcke & Markoff 2000; Yuan et al. 2002; Broderick & Loeb 2006a; Chan et al. 2009, 2015; Dexter et al. 2009; Mościbrodzka et al. 2009, 2014; Broderick et al. 2011, 2016; Dolence et al. 2012; Yuan & Narayan 2014; Gold et al. 2017; Shiokawa et al. 2017; Chael et al. 2018a).

This modeling effort is further motivated by the large amount of ancillary data that exists for EHT targets. All EHT targets are necessarily bright radio sources and thus have been the object of substantial astronomical scrutiny. Both the Galactic center (Sgr A*) and M87 have been studied across the electromagnetic spectrum, from decameter wavelengths (de Gasperin et al. 2012) to very high-energy gamma-rays ($>$1 TeV; Eckart et al. 2006). Moreover, due to the close proximity to their central black holes, both are empirically highly variable, providing statistical information about the dynamics within the millimeter-wavelength emission regions and creating opportunities to probe this dynamics directly at

multiple wavelengths (Eckart et al. 2008). Physical modeling of these sources provides the unique ability to synthesize all of these observations, which when combined with EHT data can provide a detailed description of the conditions and dynamics of material near black hole horizons.

There are substantial challenges to such a broad modeling effort. First and foremost, models of the near-horizon region are necessarily complicated, invoking multiple emission components (nonthermal and thermal emission regions with uncertain and potentially distinct locations; Özel et al. 2000; Dexter et al. 2010; Shcherbakov et al. 2012; Chandra et al. 2015; Ressler et al. 2015, 2017; Mao et al. 2017; Chael et al. 2018a; Davelaar et al. 2018), a variety of dynamical processes (orbital motion, accretion flow height, winds, jets, explosive events, etc.; Broderick & Loeb 2006a; Dexter et al. 2009; Dolence et al. 2012; Dexter & Fragile 2013; Fraga-Encinas et al. 2016; Pu et al. 2016a; Roelofs et al. 2017; Medeiros et al. 2018; Pu & Broderick 2018; Jeter et al. 2020), strong lensing in a potentially uncertain spacetime (Kerr or beyond; Bambi & Freese 2009; Johannsen & Psaltis 2010; Johannsen 2013; Broderick et al. 2014; Johannsen et al. 2016; Mizuno et al. 2018), polarization transfer effects (e.g., Faraday rotation and conversion; Huang & Shcherbakov 2011; Shcherbakov & Huang 2011; Shcherbakov et al. 2012; Mościbrodzka et al. 2017; Jiménez-Rosales & Dexter 2018), and propagation effects (e.g., interstellar scattering; Johnson & Gwinn 2015; Johnson et al. 2018; Issaoun et al. 2019). To add to the complexity of this comparison, only Fourier modes along specific tracks in the two-dimensional Fourier domain are probed via Earth–aperture synthesis on timescales that can be comparable to intrinsic source variability (Lu et al. 2016). Thus, any tools constructed to make comparisons between physical models of EHT targets and the collection of EHT and auxiliary data must be extremely flexible.

Second, there are clear emission-model independent features in many images that arise from the structure of the underlying spacetime. These include the black hole shadow—the silhouette of the black hole determined by its photon orbits, first described by Hilbert (1917) and simulated by Luminet (1979). This is bounded by the photon ring, a bright ring arising from the stacking of multiple images, in which the gross features of the spacetime are encoded. Thus, there is substantial motivation to directly extract these generic features from the EHT data alone. Again, this is complicated by the indirect relationship between the VLBI observables and the image, resulting in frequently counterintuitive conclusions. Hence, ideally, any tools for assessing the presence and properties of image structures should be able to extend to phenomenological models as well.

Third, the nature of EHT data has evolved rapidly over the past decade, growing as the sensitivity and baseline coverage improved. It is far from clear that any particular set of EHT data types is optimal for a given astrophysical or gravitational question. In some cases, new data types have been developed based on both instrumental and observational limitations (e.g., visibility polarization fractions). Similarly, intrinsic source variability has motivated the development of sophisticated statistical descriptions of observable quantities (Kim et al. 2016). Given the broad range of EHT and ancillary data types, any model comparison effort must maintain substantial flexibility in the kinds of information that it can utilize.

Fourth, there are many facets of the intrinsic data calibration that are potentially degenerate with modeling efforts. Examples include the calibration of the complex station gains and full instrument polarization response. Often these are obtained via fitting models of the instrument while assuming simplified source structures and/or iterative procedures in which the calibration is performed between attempts at source reconstruction (e.g., self-calibration). However, both of these impact the accuracy and precision of model parameter reconstructions. Thus, ideally, these calibration steps would be incorporated directly into the model-fitting effort.

Finally, in many cases, the construction of physically realistic models is computationally expensive, requiring ray tracing (relatively cheap) and radiative transfer (often expensive) through model structures. This difficulty is compounded by the often multimodal nature of the reconstructed posterior parameter distributions (see, e.g., Broderick et al. 2016). As a result, any analysis tools must be both computationally efficient and be able to exploit the large investment in high-performance computing resources.

It is to address these challenges that we have begun the development of an analysis framework for EHT and ancillary data: THEMIS. THEMIS is designed to be modular, extensible, and highly parallel, enabling the extraction of increasingly detailed information from EHT observing campaigns, both individually and in aggregate. Here we present the underlying design philosophy, structure, and validation tests of THEMIS, including the reproduction of a variety of published analyses. We then demonstrate the ability of THEMIS to extend these, presenting new analyses of phenomenological models that include the full set of published EHT observations. THEMIS has already been employed in the analysis of the first horizon-resolving imaging observations of M87 by the EHT (Papers V, VI).

In Section 2, we summarize the algorithms, components, and implementation details of THEMIS. Individual features are described in Sections 3–7. Various tests used to validate THEMIS features are presented in Section 8. A handful of novel results enabled by THEMIS are collected in Section 9. The computational performance of THEMIS and its key components, including the implications for high-performance computing (HPC) systems, is addressed in Section 10. Finally, conclusions are summarized in Section 11.

2. Summary of THEMIS

2.1. Structure

The driving motivation for THEMIS is to produce a modular and easily extensible framework for unifying existing and developing future analyses of EHT and auxiliary data. Modularity reduces the practical bars to significant contribution substantially: would-be developers need only understand the relevant elements of the interfaces between modules. In the presence of rapidly evolving data sets and model types, this is critical to ensuring the longevity of prior development efforts.

THEMIS consists of five distinct collections of components, each of which is designed to be interchangeable:

Data Structures: Management and standardization of observational data throughout THEMIS. These facilitate the rapid introduction of new data products, expand the capability of existing data products, and define the objects for which predictions are ultimately made.

Models: Any algorithm that produces a prediction for some data object given a list of parameters. Models may be physically motivated or purely phenomenological. They are directly tied to underlying data structures via the declaration of those for which predictions can be made.

Priors: Priors encapsulate preexisting information about parameter values.

Likelihoods: Likelihoods provide a method for directly comparing model and data objects. Note that in many cases elements of the underlying model may be subsumed into a likelihood (e.g., nuisance parameters that can be analytically marginalized over).

Samplers: Samplers provide methods for efficiently exploring the model parameter space, providing information about the model parameters.

In practice, there is some overlap between component classes (e.g., Models, Priors, and Likelihoods), which may be implemented in more than one way. Nevertheless, this has proven sufficiently modular to enable rapid and significant model development already.

All THEMIS-based analyses are structured in the following way:

1. Generate the desired data objects, e.g., by reading in existing data sets.
2. Create an appropriate model object, i.e., declare a model capable of making predictions for the data selected.
3. Specify prior probability distributions for each model parameter.
4. Construct the relevant likelihood objects, combining data sets as desired.
5. Execute a sampler, reporting sampler-specific parameter information (e.g., generate chains for Markov Chain Monte Carlo (MCMC) samplers).

With the execution of the analysis now conceptually modularized, variations in each stage may be made with little practical effort.

2.2. Implementation

The `main` function is kept concise and is the only element of THEMIS a user that is simply running THEMIS needs to modify. The user may choose interchangeably different EHT data set(s), theoretical model(s), priors, likelihoods, and samplers to employ. Conceptually, this function is organized in a fashion that closely follows the analysis pipeline listed at the end of the previous section to improve usability.

THEMIS also allows users to add a whole new functionality, such as additional models, which can be included easily into a clear and well-established structure. An object-oriented programming framework, along with inheritance, permits a clear and concise definition of component interfaces. Examples of how these are propagated through various THEMIS components are explicitly illustrated in the inheritance diagrams shown in Figures 1 and 2. Of practical importance, as seen in Figure 1, are the various classes of predictions enabled by a particular model class propagated to subsequent child classes (in this case, image-based models, see Section 4.1); for more details, see Section 4.

THEMIS is under version control provided by `git` with a modern, state-of-the-art branching strategy including master, development, and feature branches. Users are encouraged to generate new code branches, develop, and contribute to the code in the form of a pull request that will be reviewed by the THEMIS core development team.

A suite of tests is run regularly via a script in an effort to identify bugs or regressions as early as possible. The script performs these tests and sends a report to the THEMIS core development team. These include short tests using EHT data as well as full-scale parameter estimation validation tests similar to the ones presented in Section 8.

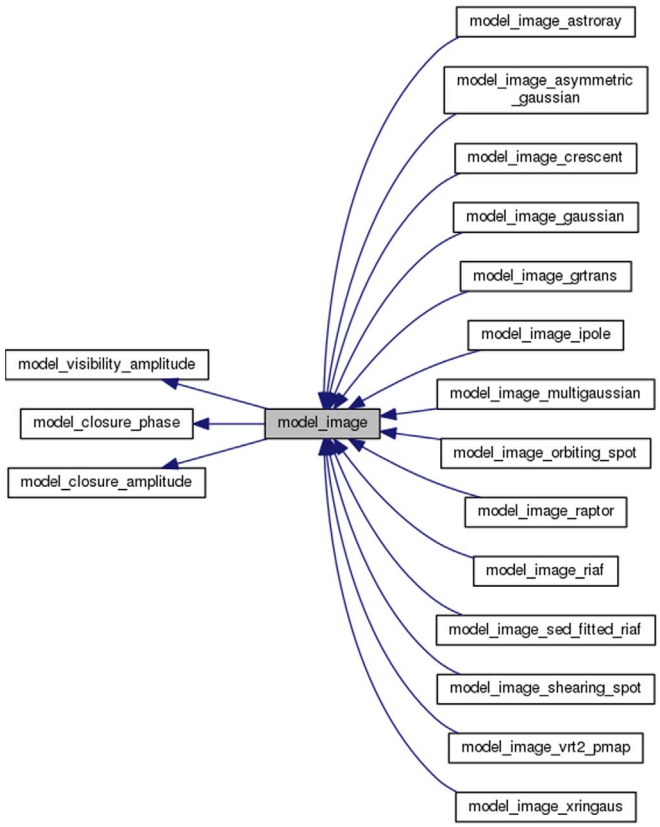

Figure 1. Inheritance diagram for the `model_image` object within THEMIS generated via `Doxygen`. These are models whose primary output is a raster image. Note that a number of models that are either analytically tractable or extend beyond a single, raster image are not shown. A full listing of THEMIS models can be found in the THEMIS documentation.

The code is written in `C++`, making it maximally portable, and has been tested on a variety of systems. THEMIS is designed with minimal dependency on external libraries to avoid installation conflicts; currently, the only required external libraries are `FFTW` (Frigo & Johnson 2005) and the Message Passing Interface (MPI).[93] Up-to-date documentation is critical in a rapid development environment. To meet this challenge, THEMIS has integrated documentation comments which may be optionally rendered via `Doxygen`[94] to produce a comprehensive, cross-linked HTML and/or PDF document.

We now turn to describing each component collection independently.

3. THEMIS Data Structures

Within THEMIS, observational data are collected in type-specific data structures. Each has a singular data element defined (a datum object) and an associated plural data structure (a data object) that provide additional input/output facilities and element access functions. At a minimum, these provide access to the values and their uncertainties. Typically, they

[93] Information on the MPI 3.1 standard can be found at www.mpi-forum.org.

[94] Information on `Doxygen` features, directives, and on how to obtain and install it may be found at www.doxygen.org.

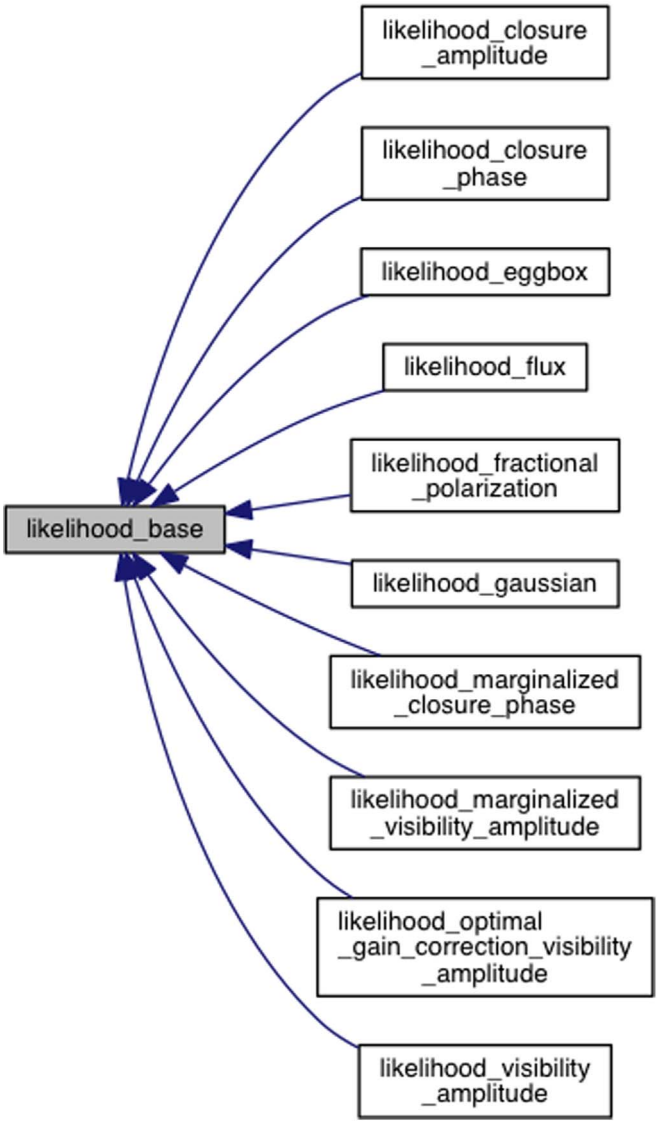

Figure 2. Inheritance diagram for the `likelihood` object within THEMIS generated via `Doxygen`.

include a variety of additional "accoutrements," i.e., information necessary or useful in modeling the data. Importantly, these accoutrements are both data-type specific and extensible: information that only becomes useful in subsequent observations or analyses can be added without modifying the data–model interface. For example, observed fluxes may initially include frequency as an accoutrement and later expand to include time, observation facility, etc.

Organizing data this way within THEMIS permits evolution in how data are employed in model comparisons and presents a simple way in which to include additional types of data that are currently unforeseen. This is especially important given the wide variety of auxiliary data that exist for EHT targets, most of which have yet to be fully utilized. This has already been implemented for a number of existing data types, including all for which EHT data have already been reported (see Table 1). We summarize each of these below.

3.1. Visibility Amplitudes

The primary product of VLBI observations is complex visibilities, corresponding to the Fourier modes of the image on the sky at spatial frequencies given by the projected baseline presented by pairs of VLBI stations. Specifically, in the absence of confounding effects, the complex visibility is given by

$$V_{ij} = \int dl dm I(l, m) e^{-2\pi i(lu+mv)}, \quad (1)$$

where (u, v) is the two-dimensional projected baseline length between the ith and jth stations expressed in units of the observed wavelength, and $I(l, m)$ is the spatial intensity distribution at angular position (l, m) (for a comprehensive introduction to radio interferometry, see, e.g., Thompson et al. 2017).[95]

In practice, these are modified by a variety of observational complications, chief among which are atmospheric absorption and phase delays at individual stations, which impact the amplitude and complex phase of V_{ij}. Of these, the latter is especially problematic, resulting in phase shifts of the V_{ij} by many times 2π, appearing to randomize the phase on every baseline. As a result, often the magnitudes of the visibilities, $|V_{ij}|$, are employed, which are subject only to a comparably modest uncertainty, 1%–20%, depending on station and atmospheric conditions (see, e.g., Johnson et al. 2015; Lu et al. 2018), albeit containing less information on the structure of the image. The number of visibility amplitudes generated by an interferometer grows quadratically with the number of stations, N, scaling as $\propto N(N-1)/2$. Throughout an

Table 1
Published Flux and EHT Data

Target	Type[a]	Obs. Campaign		N[b]	References[c]
		yr	Day(s)		
Sgr A*	**F**	**1998–2006**		**11**	**Y04, M06**
⋯	**VA**	**2007**	**100–101**	**19**	**D08**
⋯	**VA**	**2009**	**95–97**	**51**	**F11**
⋯	**CP**	**2009**	**93, 96–97**	**24**	**F16**
⋯	**CP**	**2011**	**88, 90–91,94**	**31**	**F16**
⋯	**CP**	**2012**	**81**	**25**	**F16**
⋯	**CP**	**2013**	**80–82, 85–86**	**101**	**F16**
⋯	VA	2013	80–82, 85–86	128	J15
⋯	LP	2013	80–82, 85–86	662	J15
⋯	VA	2013	80–82, 85–86	861	L18
⋯	CP	2013	80–82, 85–86	267	L18
M87	VA	2009	95–97	104	D12
⋯	VA	2012	81	56	A15
⋯	CP	2012	81	17	A15
⋯	V	2017	95, 96, 100, 101	51,119	E19

Notes.
[a] Data types include visibility amplitudes (VA), closure phases (CP), complex visibilities without phase calibration (V), interferometric linear polarization fraction (LP), and fluxes (F).
[b] Number of data points, including detections only.
[c] Y04 = Yuan et al. (2004; taken from Falcke et al. 1998; Zhao et al. 2003), M06 = Marrone (2006), D08 = Doeleman et al. (2008), F11 = Fish et al. (2011), F16 = Fish et al. (2016), J15 = Johnson et al. (2015), L18 = Lu et al. (2018), D12 = Doeleman et al. (2012), A15 = Akiyama et al. (2015), E19 = Paper III. The data set listed in bold is used for reproducing previous results as part of the validation tests (see Section 6.3).

[95] Note that the default definition within the EHT following 2017 returns the complex conjugate of the complex visibilities, i.e., the sign in the exponent of Equation (1) is positive. Within THEMIS, this is corrected at the data input/output stage.

observing campaign, the rotation of Earth produces a large number of independent measurements at different projected baselines.

A large number of EHT visibility amplitudes have already been published for the primary EHT targets, including Sgr A* (Doeleman et al. 2008; Fish et al. 2011; Johnson et al. 2015; Lu et al. 2018) and M87 (Doeleman et al. 2012; Akiyama et al. 2015). We list these in Table 1.

3.2. Closure Phases

The atmospheric phase delays introduce random station-specific phase errors that complicate the reconstruction of the phase of the complex visibilities. While strategies exist to model the phase errors, including self-calibration and phase referencing,[96] a simple and efficient way in which to obtain some information about these phases is via the closure phase,

$$\Phi_{ijk} = \arg\left(V_{ij} V_{jk} V_{ki}\right), \tag{2}$$

i.e., the sum of the phases of a triplet of visibilities measured on the baselines between some triplet of stations.[97] Because the baselines "close," i.e., $(u, v)_{ij} + (u, v)_{jk} + (u, v)_{ki} = 0$, all station-specific phase errors vanish identically, leaving a quantity that is subject only to nonclosing errors and dominated by the image structure (Paper III). Of particular importance is the fact that closure phases are also insensitive to the image blurring induced by the diffractive component of the interstellar scattering. Closure phases are not unique—for an array with N stations only $(N-1)(N-2)/2$ are independent—a result that is presaged by their independence from the phase delays.

Closure phases have been reported for a number of years by the EHT for Sgr A* in Fish et al. (2011) and Lu et al. (2018), and for M87 in Akiyama et al. (2015), as summarized in Table 1. More recently, a large set of closure phases has been reported for the 2017 April EHT observations of M87 (Paper III).

3.3. Closure Amplitudes

Station-specific amplitude calibration errors can also be mitigated by combining visibilities measured on multiple baselines. The closure amplitude is constructed from combinations of visibilities measured on four stations,

$$\mathcal{V}_{ijkm} = \frac{|V_{ij}||V_{km}|}{|V_{ik}||V_{jm}|}, \tag{3}$$

and is insensitive to variations in the flux calibration and phase delays. Again, closure amplitudes are also insensitive to the image blurring induced by the diffractive component of the interstellar scattering. As with the closure phase, this comes at the price of uniqueness; there are only $N(N-3)/2$ independent closure amplitudes.

Closure amplitudes constructed from EHT data have not yet been published, primarily due to the limited number of stations participating in early observations. However, recent observations have generated a number of trivial closure amplitudes, i.e., amplitudes for which one baseline is very short ($\lesssim$10 km), as part of the calibration process (see, e.g., Johnson et al. 2015). More recently, a large set of closure amplitudes have been reported for the 2017 April EHT observations of M87 (Paper III).

3.4. Interferometric Polarization Fractions

All EHT sites observe simultaneously in two polarization states (Paper II). From these, it is possible to construct complex visibilities in all four Stokes parameters, (I, Q, U, V) (see, e.g., Johnson et al. 2015). Independently, these can be used to construct visibility amplitudes, closure phases, and closure amplitudes. However, additional information may be obtained by combining observations made in different Stokes parameters. The interferometric polarization fraction,

$$\breve{m}_{ij} = \frac{\sqrt{|V_{ij}^{Q}|^2 + |V_{ij}^{U}|^2}}{|V_{ij}|}, \tag{4}$$

where $V_{ij}^{Q,U}$ are the visibilities associated with Stokes Q and U, and V_{ij} is the visibility defined in Equation (1), is the extension of the familiar polarization fraction to the individual Fourier modes of the image. $\breve{m}$ is not to be mistaken with the Fourier transform of the linear polarization fraction as measured in the image domain. Unlike the standard polarization fraction, $\breve{m}_{ij}$ may be larger than unity and can exhibit counterintuitive pathologies for even simple source models (see the discussion surrounding Figure S6 in the supplemental material in Johnson et al. 2015). If the circular polarization vanishes, like closure amplitudes, the interferometric polarization fractions are insensitive to station-specific flux calibration uncertainties and the diffractive component of the interstellar scattering. This approximation is frequently quite good: for the astronomical sources of primary interest to the EHT, the integrated circular polarization fraction is typically $\lesssim$1% (see, e.g., Muñoz et al. 2012; Bower et al. 2018).

Interferometric polarization fractions have been reported for Sgr A* and indicate the presence of ordered horizon-scale polarization structures (Johnson et al. 2015). We summarize these in Table 1.

3.5. Flux Measurements

A key auxiliary set of observations is the spectral energy density distributions (SED) for primary EHT targets, which typically place strong limits on the emitting particle distributions, which are otherwise uncertain. In addition, multi-wavelength light curves are a key probe of the nature and origin of variability in the emission regions of the source. Both empirical constraints are intrinsically encoded in measurements of the unresolved source flux, F_ν, effectively equivalent to the visibility amplitudes measured at "zero baseline," i.e., a single dish. The distinction between these arises in the accoutrements associated with the data, e.g., the origin of the observation, wavelength, time, etc.

Multiple sets of flux measurement data for Sgr A* and M87 exist. For Sgr A*, one set is summarized in Table 1.

4. THEMIS Models

Within THEMIS, a model is any algorithm capable of generating a prediction for any THEMIS data type. Thus,

[96] These may include extragalactic background sources (see, e.g., Broderick et al. 2011; Thompson et al. 2017), simultaneous observations at other wavelengths or in alternative Stokes parameters (Middelberg et al. 2005; Johnson et al. 2014), or strong spectral lines (Herrnstein et al. 1998; Reid 1999).

[97] This is the argument of the bispectrum.

THEMIS models are closely aligned with THEMIS data structures—for each data type, there is a corresponding base model type. Models can be based on multiple base model types, i.e., they are capable of generating predictions for more than one type of data. This enables a broad, easily extensible, and backwards compatible framework for defining models that permits incrementally increasing sophistication. Importantly, it provides a uniform interface for both phenomenological models, which are designed to make predictions for a handful of data types, and physically motivated models, which can simultaneously make predictions for a wide variety of data types.

The manner in which predictions are made is not prescribed. That is, where analytical expressions for the relevant data type exist (e.g., visibilities from simple geometric models), models are capable of employing these. For more complex models, numerical computations are often required. In anticipation of numerically produced predictions, THEMIS permits the passing of an accuracy parameter for each value that specifies the accuracy with which these must be generated; typically, setting this to 25% of the measurement uncertainty is sufficient to generate accurate parameter estimates (see Appendix A).

4.1. Image-based Models

Because the EHT directly probes the structure of horizon-scale images, THEMIS contains an image-based model type; how this depends on the underlying data-based model types and some examples are shown in Figure 1. This provides a set of utilities for generating and manipulating visibility-based data from models that primarily generate images.

Because image generation is frequently computationally intensive, the image-based model introduces an additional position angle parameter, permitting the specific model implementations to dispense with trivial image rotations, leading to a substantial potential reduction in the time required to sample a broad range of parameters.

Once generated, images are padded with zeros by a factor of 8 by default to sinc-interpolate in the numerically computed complex visibilities. The complex visibilities are computed on a two-dimensional grid of (u, v) values via a two-dimensional fast Fourier transform (FFT) using the FFTW library (Frigo & Johnson 2005). There are no restrictions on the image dimensions, though it is expected that the image is computed on a rectilinear grid with uniform pixel size; dimensions that factor into small primes will be marginally faster.

Complex visibilities are then estimated at arbitrary (u, v) via interpolation. By default, THEMIS employs bicubic interpolation, though a user may specify bicubic spline interpolation if desired. From these, the closure phases are constructed via Equation (2). While visibility magnitudes may also be constructed from the interpolated complex visibilities, it is considerably more accurate to interpolate the visibility magnitudes directly.[98] These are then used directly or to compute closure amplitudes via Equation (3). The accuracy of this interpolation depends on the degree to which image features are captured by the image resolution and field of view. Ultimately, this must be empirically assessed on a model-by-model basis. For the compact radio sources like those anticipated to be relevant for EHT targets, these interpolation errors are typically a significantly subdominant source of uncertainty in model evaluation and parameter estimation.

[98] The magnitude of the gradient of the complex visibility and the visibility amplitude are related via $|\nabla V|^2 = (\nabla|V|)^2 + |V|^2(\nabla\phi)^2 \geqslant (\nabla|V|)^2$, and thus the former is generally smaller than the latter. As a result, the errors in interpolation at any order are typically smaller when interpolating visibility amplitudes directly. Alternatively, this permits considerably smaller image sizes when only amplitudes are required.

4.2. Phenomenological Geometric Models

Within THEMIS, a number of phenomenological geometric models have been implemented. These are models for which no underlying physical emission mechanism is identified for the origin of the image structures. However, such models are capable of extracting signatures of geometric features associated with underlying physical processes of interest, e.g., black hole shadows. Currently implemented phenomenological models include the following.

4.2.1. Symmetric Gaussian

Historically, the first shadow size estimates from millimeter-VLBI observations of Sgr A* and M87 arose from fitting symmetric Gaussians to visibility amplitude measurements (Doeleman et al. 2008). Therefore, we have implemented within THEMIS a model consisting of a single symmetric Gaussian component, characterized by a size, σ; and an amplitude, V_0. This makes predictions for visibility amplitudes, closure phases (trivially zero), and closure amplitudes.

4.2.2. Asymmetric Gaussian

The introduction of asymmetry in millimeter-VLBI images was initially characterized by an asymmetric Gaussian. Within THEMIS, we have implemented such a Gaussian model parameterized as in Broderick et al. (2011), and characterized by a size, σ; an asymmetry parameter, A; the amplitude, V_0; and the position angle, ξ.

4.2.3. Multiple Symmetric Gaussians

THEMIS also includes a model consisting of an arbitrary number of symmetric Gaussian components, each characterized by a size, σ_j; location, (x_j, y_j); and amplitude, V_j.

4.2.4. Crescent Model

THEMIS includes an implementation of the crescent model described in Kamruddin & Dexter (2013), for which the image is obtained by subtracting two nonconcentric disks, with the smaller disk lying completely inside the larger one. The complex visibilities for this model can be obtained analytically and are given by Equation (3) of Kamruddin & Dexter (2013). As in Kamruddin & Dexter (2013), we reparameterize this in terms of an amplitude, V_0; overall size, R; relative thickness, ψ; degree of symmetry, τ; and the position angle, ξ. Both ψ and τ are defined on the unit interval.

4.2.5. The "Xsringauss" Model

THEMIS also contains an implementation of the nine-parameter xsringauss model proposed in Benkevitch et al. (2016). This model was constructed in an effort to mimic a more realistic black hole accretion image like the ones commonly obtained from physically motivated models. The xringaus image is the combination of an eccentric slashed ring (i.e., with a linear brightness gradient) and an elliptical Gaussian located in the thicker side of the ring.

This model is then described by nine parameters: the zero-spacing flux, V_0; the external radius, $R_{\rm ex}$; the internal radius, $R_{\rm in}$; the distance between the centers of the circles, d; the "fading" parameter controlling the minimum brightness; the Gaussian axis sizes, a and b; the fraction of the total flux in the Gaussian, g_q; and the position angle, ξ. The complex visibilities for this model, in terms of these parameters, can be also obtained analytically. The reader is referred to Section 2 of Benkevitch et al. (2016) for a more detailed description.

4.2.6. Visual Binary

THEMIS also features a model of two radio emitting symmetric Gaussian components in orbit around each other. The model is characterized by 13 parameters including the total flux F_i, size σ_i, and spectral index α_i of each component; the total mass of the system M, the binary mass ratio $q \leqslant 1$, the orbital separation R, the source distance d, the phase offset Φ_0, the cosine of the inclination angle $\cos(i)$ of the orbital angular momentum vector, and the position angle in the sky ξ. This model includes (and therefore also takes advantage of) relativistic effects such as Doppler boost and relativistic aberration. It is explicitly time dependent while being fully analytic and thus fast to evaluate.

This model is to be compared to long-timescale monitoring campaigns of sources such as OJ 287 or other binary candidates. Details will be published in a separate paper that focuses on this topic.

4.3. Interstellar Scattering Models

Interstellar scattering modifies the intrinsic images of Sgr A* by both blurring the image (diffractive component) and adding small-scale structures associated with a random realization of refractive modes that vary slowly throughout the night (refractive component; see, e.g., Johnson & Narayan 2016). These significantly modify visibilities on long baselines and must be included in analyses of EHT observations of Sgr A*.

In the ensemble-average limit, i.e., when many realizations of the fluctuations in the scattering screen may be averaged over, only the diffractive component is relevant. This appears as an image smoothing via convolution with a Gaussian kernel whose parameters depend on the details of the intervening scattering screen(s). THEMIS has implemented two models for addressing interstellar scattering, both in the ensemble-average limit, which we list below. In both, the impact of scattering is imposed directly on the visibilities, for which the convolution in image space reduces to a multiplicative factor. Within THEMIS, each is implemented as a model that modifies an existing intrinsic model, with the latter introducing additional parameters. Hence, scattering provides an explicit example of how the modular structure of THEMIS enables the rapid construction of new models.

While only a very simple set of scattering models has been implemented in THEMIS thus far, more complex models are available in the literature. Physically motivated models that exhibit a smooth transition from a quadratic to a general power-law wavelength dependence for the size of the scattering kernel may be found in Psaltis et al. (2018). In these, the wavelength at which the transition occurs is determined by the underlying physical parameters of the screen. Updated values of the scattering kernel size from long-wavelength measurements within the context of these models may be found in Johnson et al. (2018). Implementing these updated models is left for future development.

4.3.1. Default Diffractive Screen

Multiwavelength observations have produced a model for the scattering kernel that is asymmetric and wavelength dependent, consistent with that anticipated by models of the scattering screen that invoke Kolomogorov turbulence within a plasma sheet (Bower et al. 2006; Johnson et al. 2018; Psaltis et al. 2018). The associated semimajor and semiminor axis sizes are given by

$$\sigma_{\rm maj} = 9.39 \left(\frac{\lambda}{1.3\ {\rm mm}}\right)^2 \mu{\rm as}$$
$$\sigma_{\rm min} = 4.59 \left(\frac{\lambda}{1.3\ {\rm mm}}\right)^2 \mu{\rm as} \quad (5)$$

and are oriented such that the major axis lies along the position angle $\xi = 78^\circ$ east of north.

4.3.2. Parameterized Diffractive Screen

Recently, it has been shown that even for thin scattering screens, the wavelength dependence of anisotropic scattering screens may be substantially more complicated (Psaltis et al. 2018). The main uncertainty is the inner scale of the turbulence within the screen, corresponding to the dissipative scale within the sheet. For some plausible values, the wavelength dependence could depart from that found in Bower et al. (2006) near 1.3 mm. As a result, a second scattering model has been implemented in which $\sigma_{\rm maj}$, $\sigma_{\rm min}$, and the position angle are all parameterized as power laws of wavelength with unknown coefficients and powers. That is,

$$\sigma_{\rm maj} = \sigma_{\rm A}\left(\frac{\lambda}{\lambda_p}\right)^{\alpha}, \quad \sigma_{\rm min} = \sigma_{\rm B}\left(\frac{\lambda}{\lambda_p}\right)^{\beta},$$
$$\text{and} \quad \xi = \xi_0 + \xi_1\left[\left(\frac{\lambda}{\lambda_p}\right)^{\gamma} - 1\right], \quad (6)$$

where the seven parameters, $\sigma_{\rm A}$, $\sigma_{\rm B}$, ξ_0, ξ_1, α, β, and γ, may be varied. The pivot wavelength, λ_p, is not a parameter (being fully degenerate with $\sigma_{\rm A,B}$ and set by the user to a convenient value.

4.4. Native Physical Models

The past two decades have seen the development of a number of physically motivated models that employ ray tracing and radiative transfer in black hole spacetimes. These have two main components: the construction of photon trajectories within the spacetime under consideration and the radiative transfer through some emitting plasma distribution. Both elements are directly affected by variations in the spacetime structure, with the emission also depending on a number of astrophysical considerations.

While this class of models is substantially more complicated than geometric models, their physical origin presents a number of significant advantages. First, they are capable of making predictions for a wide range of observations, making it possible to bring far more empirical data to bear upon them. For example, they necessarily make simultaneous, self-consistent

predictions for images, fluxes, variability, and polarization features of the EHT and auxiliary data (Bromley et al. 2001; Broderick & Loeb 2006b; Dexter et al. 2009; Huang et al. 2009; Mościbrodzka et al. 2014; Chan et al. 2015; Pu et al. 2016a; Gold et al. 2017; Chael et al. 2018a). Hence, physical modeling enables a concordance fitting effort that promises far more power to constrain the nature of the emission region (Chan et al. 2015; Broderick et al. 2016; Gold et al. 2017). Second, the spacetime structure impacts the image in many ways beyond gravitational lensing. The dynamics of the material in the emission region modifies its optical depth and therefore appearance (Broderick & Blandford 2003, 2004; Broderick & Loeb 2005, 2006a, 2009b; Dexter et al. 2009; Mościbrodzka et al. 2009; Pu et al. 2016a; Gold et al. 2017; Bronzwaer et al. 2018; Chan et al. 2018; Mościbrodzka & Gammie 2018; Jeter et al. 2020). Thus, in principle, modeling the brightness distribution offers additional probes of gravity (Broderick & Loeb 2006a; Johannsen 2013; Broderick et al. 2014; Johannsen et al. 2016; Mizuno et al. 2018). Third, it provides direct information about the high-energy astrophysical processes responsible for the growth of black holes and the launching of jets (Broderick & Loeb 2009b; Dexter et al. 2009; Levinson & Rieger 2011; Mościbrodzka et al. 2014; Broderick & Tchekhovskoy 2015; Hirotani & Pu 2016; Gold et al. 2017).

Within THEMIS, two general relativistic ray-tracing and radiative transfer packages are provided. The first of these is the vacuum ray-tracing and radiative transfer package Vacuum Ray Tracing and Radiative Transfer (VRT^2). VRT^2 is based on the plasma radiative transfer package described in Broderick & Blandford (2003, 2004) and provides a modular framework for adding novel plasma distributions, radiative transfer mechanism, and spacetime structures. It was the basis for the images generated in, e.g., Broderick et al. (2011) and used in the analysis of Broderick et al. (2016). It also natively interfaces with THEMIS, having been written in the same programming language (C++), in a similar style. Models based on VRT^2 within THEMIS include those listed below.

In addition, the vacuum ray-tracing and radiative transfer package Odyssey described in Pu et al. (2016b) has also been incorporated within THEMIS. Based on the ray-tracing algorithm in Fuerst & Wu (2004) and the radiative transfer formula presented in Younsi et al. (2012), Odyssey can exploit graphics processing unit (GPU) cards to realize substantial speed gains for models that employ it. It requires the compute unified device architecture (CUDA)-enabled GPU cards and the CUDA compiler nvcc. Again, like THEMIS, Odyssey is implemented in C/C++ and CUDA C/C++, making its integration straightforward.

4.4.1. SED-fitted RIAF

This is an image at a single wavelength associated with the radiatively inefficient accretion flow (RIAF) models described in Broderick & Loeb (2006b) and refined in Broderick et al. (2011). This model employs a tabulated set of accretion flow parameters, obtained at different black hole spins and inclinations, that reproduce the observed compact radio SED of Sgr A*. The model parameters are the dimensionless spin magnitude, a (in the range [0, 1]); the cosine of the inclination, $\cos\theta$, ([−1, 1]); and the position angle, ξ ([−180°, 180°], as part of a model image). The accretion flow angular momentum is assumed to be aligned with the black hole spin. The intensity normalization may be included via the likelihood (see Section 6.6). An example image from the THEMIS-integrated VRT^2 package is shown in Figure 3.

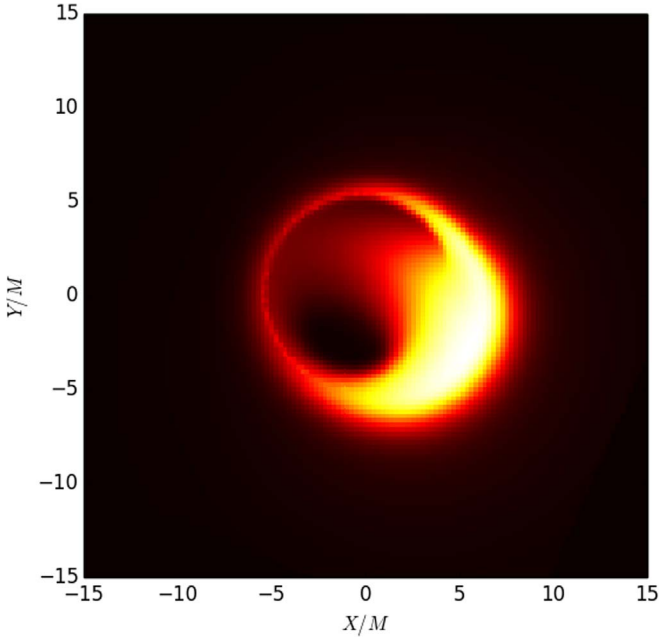

Figure 3. Image produced by the radiative transfer module VRT^2 showing the resulting best-fit RIAF model after sampling the full parameter space. This reproduces the result in Broderick et al. (2016). The X- and Y-axes show image coordinates in units of the gravitational radius.

4.4.2. Extended RIAF

This is an extension of the SED-fitted RIAF model that permits a wide range of structural parameters in the RIAF model to vary. This consists of two populations of synchrotron-emitting electrons, orbiting a Kerr black hole in the presence of a toroidal magnetic field. Specifically, the proper number density and temperature of a thermal population of electrons are given by

$$n_{\rm th} = n_{e,t} r^{\eta_t} e^{-z^2/2h_t^2R^2}, \quad T_{\rm th} = T_e r^{\tau_t}, \tag{7}$$

where $z = r\cos\theta$ and $R = r\sin\vartheta$, where r is the standard Boyer–Lindquist radius (measured in GM/c^2) and ϑ is the Boyer–Lindquist polar angle. Similarly, the proper number density of the nonthermal electrons is given by

$$n_{\rm nth} = n_{e,\rm nt} r^{\eta_{\rm nt}} e^{-z^2/2h_{\rm nt}^2R^2}, \tag{8}$$

and has a power-law distribution in microscopic Lorentz factor above $\gamma_{\rm min}$ with a power law corresponding to an optically thin spectral index of α (i.e., $2\alpha - 1$). These are emitting within a toroidal magnetic field with comoving strength

$$\frac{B^2}{8\pi} = \beta^{-1}\frac{m_p c^2}{6r}, \tag{9}$$

and orbiting with a four-velocity outside of the innermost stable circular orbit (ISCO) given by

$$u_\mu = u_t(1, 0, 0, \kappa\ell_K), \tag{10}$$

where ℓ_K is the specific Keplerian angular momentum and u_t is determined by the standard normalization condition on u^μ; inside of the ISCO, the material plunges on ballistic orbits

(Cunningham 1975). Thus, there are 15 parameters: black hole spin, a; cosine of the black hole spin inclination, $\cos\theta$; black hole spin position angle, ξ; thermal electron density normalization, $n_{e,t}$, radial power law, η_t, and scale height, h_t; electron temperature normalization, T_e; radial power law, τ_t; nonthermal electron density normalization, $n_{e,\rm nt}$, radial power law, $\eta_{\rm nt}$, and scale height $h_{\rm nt}$; minimum microscopic Lorentz factor, $\gamma_{\rm min}$, and spectral index, α; plasma beta, β, and sub-Keplerian fraction κ. Note that subsets of these may be held fixed or varied simultaneously via the definition of a wrapper model.

4.4.3. Orbiting Hot Spots

Major dissipative events within the accretion flow, such as magnetic reconnection events and shocks, can generate initially compact, orbiting, synchrotron-emitting hot spots. These may increase the emission of Sgr A* by orders of magnitude before inducing any dynamical effects. Therefore, they may be roughly modeled as orbiting, Gaussian nonthermal particle overdensities that subsequently synchrotron emit in the radio and infrared, restricted to the equatorial plane (Broderick & Loeb 2005, 2006a). To model the velocity profile of the spot we use a two-parameter, $(\alpha_r, \kappa) \in [0, 1]$, four-velocity given by

$$u^\mu = (u^t, u^r_{\alpha_r}, 0, u^t\Omega_\kappa). \tag{11}$$

Here, $u^r_{\alpha_r} = u^r_{\rm K} + \alpha_r(u^r_{\rm ff} - u^r_{\rm K})$ and $\Omega_\kappa = \Omega_{\rm ff} + \kappa(\Omega_{\rm K} - \Omega_{\rm ff})$, where the $_{\rm K,ff}$ subscripts denote Keplerian and freefall motion respectively, and $\Omega_i = u^\phi_i/u^t_i$ (see also Pu et al. 2016a). Equation (11) is a two-parameter description of the flow dynamics and can also be applied to the extended RIAF model in Section 4.4.2, thereby generalizing Equation (10). Thus, there are 10 parameters needed for this model: black hole spin, a; cosine of the black hole spin inclination $\cos\theta$; black hole spin position angle, ξ; central spot nonthermal electron density $n_{e,\rm spot}$; spot radial size R_s; initial spot location in time, t_0, radius, r_0, and azimuthal angle ϕ_0; the sub-Keplerian parameter, κ; and the radial infall parameter α_r.

4.4.4. Shearing Hot Spots

In practice, hot spots will subsequently shear and cool. Thus, THEMIS also includes a shearing hot spot model (Jeter et al. 2020) that incorporates the expansion of the hot spots within a background accretion flow. The parameters of this model are identical to the orbiting spot model above.

4.5. External Physical Models

There is no intrinsic bar to including additional ray-tracing and radiative transfer packages within THEMIS. Doing so offers a number of benefits, including the ability to rapidly generate new models within THEMIS itself, efficient parallelization, and improved portability. However, native integration is not necessary. It is often initially faster, and occasionally necessary, to externally include modeling software. For THEMIS, this has been done for a number of existing packages:

GRTRANS: A publicly available general relativistic, polarized radiative transfer code written in FORTRAN, described in Dexter (2016) and Dexter & Agol (2009). GRTRANS and by extension also THEMIS are coupled to the HARM3D general relativistic, magnetohydrodynamic (GRMHD) code (Gammie et al. 2003; Noble et al. 2006; Dexter et al. 2009; McKinney et al. 2014).

ASTRORAY: A significantly extended version of the general relativistic polarized radiative transfer code written in C/C++ based on Shcherbakov (2014) and substantially extended in Gold et al. (2017). ASTRORAY and by extension THEMIS are coupled to HARM3D (Gammie et al. 2003; McKinney et al. 2012, 2014).

iPOLE: A publicly available general relativistic, polarized radiative transfer code described in Mościbrodzka & Gammie (2018) based on the covariant formulation presented in Gammie & Leung (2012) and written in standard C. iPOLE and by extension THEMIS are coupled to HARM3D (Gammie et al. 2003; Dolence et al. 2009; Mościbrodzka et al. 2009).

RAPTOR: A publicly available general relativistic radiative transfer code described in Bronzwaer et al. (2018) written in standard C. RAPTOR and by extension THEMIS are coupled to the BHAC (Porth et al. 2017; Olivares et al. 2018) GRMHD code, HARM3D (Gammie et al. 2003; Mościbrodzka et al. 2009) and is GPU capable.

BHOSS: A publicly available general relativistic radiative transfer code described in Z. Younsi et al. (2019, in preparation) written in Fortran 95/2003 (see also Fuerst & Wu 2004; Younsi et al. 2012, 2016). BHOSS and by extension THEMIS are coupled to the BHAC (Porth et al. 2017; Olivares et al. 2018), HARM3D GRMHD code (Gammie et al. 2003; Dolence et al. 2009; Mościbrodzka et al. 2009) and H-AMR (Liska et al. 2018; Chatterjee et al. 2019).

Note that many of these are directly coupled to a variety of existing GRMHD simulation codes such as HARM3D and BHAC. As of now, THEMIS has successfully interfaced, in at least a limited form, with the vast majority of the image-generation tools employed by the EHT collaboration.

5. Priors

Preexisting information about parameters may be imposed via priors. THEMIS provides a number of potential priors for individual parameters. These may be imposed in two distinct ways: as "priors" that modify the likelihood and "transforms" that modify the parameter values. Within THEMIS, "priors" add a term associated with a given prior distribution. These are most easily implemented and understood. However, they can be inefficient, assuming that the sampler will efficiently incorporate the modified likelihood. In contrast, "transforms" impose priors indirectly by mapping the variable being sampled into the desired prior via a coordinate transformation. These are more complicated to implement, typically requiring the integration of the desired prior probability distribution. However, they are optimally efficient, permitting the sampler to apply a more natural distribution. Note that "transforms" may be implemented intrinsically within models by choosing a convenient set of parameters.

Likelihood evaluation is short-circuited on the evaluation of priors, i.e., where the prior has zero probability (e.g., outside the limits of a linear range), the likelihood is not evaluated but rather returns the appropriate vanishing value. This achieves two goals: first, THEMIS is made marginally more efficient by avoiding unnecessary computation, and second, permits priors

to be used to avoid unphysical parameter combinations, where models may return nonsensical results, e.g., negative densities passed to a radiative transfer code or black hole spin outside the range permitted by general relativity.

Currently, THEMIS has only implemented priors and transforms of a single variable. This is sufficient for most situations. However, there are situations that may benefit from priors that depend on many parameters, e.g., enforcing an ordering among the intensities of multiple Gaussian components, thereby eliminating the trivial degeneracy associated with swapping components. Nevertheless, there is no reason that such a prior cannot be implemented within THEMIS.

Implemented priors include:

1. *None:* a flat prior without a boundary.
2. *Linear:* a flat prior given two bounding values.
3. *Logarithmic:* a logarithmic prior given two bounding values.
4. *Gaussian:* a Gaussian prior given a mean and standard deviation.

And implemented transforms include:

1. *None:* no transformation (default).
2. *Fixed:* returns a single, user-defined value.
3. *Logarithmic:* effectively imposes a logarithmic prior.

6. Likelihoods

Models and data are systematically compared via likelihoods, which express the probability that the data was obtained from the model. Within THEMIS, a likelihood is any method for taking a parameter vector, $\boldsymbol{p}$, and constructing a log-likelihood, $\mathcal{L}$. When this is generated using a THEMIS data object (consisting of a number of individual values) and a THEMIS model object, the log-likelihood is the probability of obtaining the data given the model. Likelihoods can be combined with user-supplied weights, enabling the combination of various data sets. However, when doing so it is assumed that the model parameters are unchanged, i.e., the same set of model parameters are to be supplied to each likelihood being used. All likelihoods expect a matching data type and model type, e.g., visibility amplitude data and a model that generates visibility amplitude predictions.

The likelihood generally requires information about the underlying error distribution of the data, which is typically provided via an error estimate. It is not required within THEMIS to assume Gaussian errors, i.e., likelihood classes that assume alternative error distributions (e.g., Rice distributions, etc.) are possible. In some instances, this flexibility is important, e.g., for quantities constructed from quotients (closure amplitudes, Section 6.4, and interferometric polarization fractions, Section 6.5). It is, however, often convenient numerically to presume Gaussian errors when permissible, enabling analytical simplifications that greatly improve the efficiency of THEMIS (Sections 6.6 and 6.7). Similarly, all currently implemented likelihoods assume the data values are independent—this, too, may be relaxed in principle. An obvious example of both that is of considerable interest is the covariance induced by the refractive modes in the scattering screen, the implementation of which is left for future development.

Likelihoods also can incorporate model features. In many instances, a subset of model parameters may be analytically marginalized over and in the process subsumed into the likelihood itself. We have implemented a number of examples of such "marginalized" likelihoods, i.e., likelihoods in which sets of nuisance parameters have been treated analytically. It is natural to include key systematic uncertainties of the EHT, e.g., the structure of the refractive scattering screen, in this fashion, though this is left for future development.

The likelihoods currently implemented in THEMIS include the following.

6.1. Test Cases

To facilitate testing samplers, THEMIS includes five artificial likelihoods with given distributions. The first is a multi-dimensional Gaussian, with user-specified mean and size, see Section 8.1.3. The second, the egg box, is considerably more complicated, producing a highly multimodal likelihood function in five dimensions:

$$\mathcal{L}(\boldsymbol{p}) = [2 + \prod_{i=0}^{4} \cos(p_i)]^5 . \tag{12}$$

The number of peaks can be set by the range over which the priors permit the parameters, $\boldsymbol{p}$, to vary. This is typically used to assess the ability of a sampler to accurately find widely separated, high-likelihood regions. The results of the egg box are presented in Section 8.1.2.

The third artificial likelihood is the 2D Rosenbrock function:

$$\mathcal{L}(\boldsymbol{p}) = 100(x_1 - x_0^2)^2 + (1 - x_0)^2. \tag{13}$$

The test results are presented in Section 8.1.4.

The fourth artificial likelihood is the five-dimensional Griewank problem:

$$\mathcal{L}(\boldsymbol{p}) = \left[1 + \frac{1}{4000}\sum_{i=0}^{4} p_i^2 - \prod_{i=0}^{4} \cos(p_i)/\sqrt{i+1}\right]. \tag{14}$$

It features an egg box component and a long-tailed parabolic contribution. The test results are presented in Section 8.1.5.

The fifth artificial likelihood is a three-dimensional Cauchy/Lorentz distribution:

$$\mathcal{L}(\boldsymbol{p}) = -\frac{d+1}{2}\log\left[\Sigma_{i=1}^{d}(x_i - m)^2 + \alpha^2\right], \tag{15}$$

where $d = 3$ is the chosen dimension and $m = 0.1$ the width of the line profile. It features heavy and long tails and has no well-defined mean, and therefore makes for a challenging sampling problem. The test results are presented in Section 8.1.6.

6.2. Visibility Amplitudes

THEMIS includes a log-likelihood that assumes Gaussian errors for visibility amplitudes:

$$\mathcal{L}(\boldsymbol{p}) = -\sum_j \frac{[|V|_j - |\hat{V}|_j(\boldsymbol{p})]^2}{2\sigma_j^2}, \tag{16}$$

where $|V|_j$ and σ_j are the observed visibility amplitudes and their errors, and $|\hat{V}|_j(\boldsymbol{p})$ are the model visibility amplitudes given parameters $\boldsymbol{p}$. Note that the true visibility amplitude error distribution is given by the Rice distribution, and for low signal-to-noise ratio (S/N) is both biased and non-Gaussian (Thompson et al. 2017). However, when data are selected such that $\mathrm{S/N} \geqslant 2$ and approximately debiased via $|V|_j \to \sqrt{|V|_j^2 - \sigma_j^2}$, the visibility amplitude error distribution

is within 8% of an unbiased Gaussian distribution at all $|V|$, and reproduces the mode to better than 1% and the 68% and 95% cumulative widths to better than 6% (see Appendix B). Currently, the user is expected to independently implement the debiasing procedure and S/N cut in the generation of the data tables prior to reading them in THEMIS. These were made for the THEMIS analyses in Papers V and VI.

6.3. Closure Phases

In principle, closure-phase errors are non-Gaussian and correlated among station triangles. However, this is simplified in practice for EHT observations by two facts. First, current and future EHT data frequently have high S/Ns. For closure phases constructed from complex visibility measurements with S/N > 2, the closure-phase errors are well approximated by Gaussians (Rogers et al. 1984; Ricarte & Dexter 2015). Second, existing and anticipated EHT observations are performed with a highly heterogeneous array, in which a nondegenerate set of closure phases may be selected such that a single station dominates the sensitivity of each triangle. In this case, the off-diagonal elements of the covariance become subdominant, and the closure-phase measurements effectively independent (Blackburn et al. 2020).

Therefore, THEMIS includes a log-likelihood that assumes Gaussian errors for closure phases:

$$\mathcal{L}(\boldsymbol{p}) = -\sum_j \frac{\Delta^2(\Phi_j - \hat{\Phi}_j(\boldsymbol{p}))}{2\sigma_j^2}, \tag{17}$$

where Φ_j and σ_j are the observed closure phases and their errors, $\hat{\Phi}_j(\boldsymbol{p})$ is the model closure phase given parameters $\boldsymbol{p}$, and $\Delta(x)$ is the angular difference in the range $[-180°, 180°)$. This is similar to, but distinct from, the visibility amplitude likelihood in that the difference selects the branch that minimizes the angular difference. In the limit of small σ_j, this is identical to the circular dispersion in Medeiros et al. (2017) and the circular statistics in Chael et al. (2018b); at large σ_j, all of these approximations differ, though in this limit the closure phases become uninformative. We adopt this simpler prescription as it facilitates addressing scattering-induced closure-phase fluctuations later.

Again, the user is expected to independently implement the S/N cut in the generation of the data tables prior to reading them in THEMIS. These were made for the THEMIS analyses in Papers V and VI.

6.4. Closure Amplitudes

Closure amplitudes provide an example of a non-Gaussian likelihood within THEMIS. Because closure amplitudes are constructed via taking ratios of visibility amplitudes, the likelihood of a single value exhibits a significant asymmetry and extended tail toward large values, characteristic of quotient distributions (see Appendix B). For S/N $\geqslant$ 4, this is well approximated by a Gaussian quotient distribution,[99] given in Equation (B17). In principle, this can make use of ancillary information in the form of station system equivalent flux densities (SEFDs), though this is left for future development. Thus, at present, we assume that the parameter ρ, defined in Equation (B14), is fixed to unity, for which the Gaussian quotient approximation is accurate at all $\mathcal{V}$ to better than 13% for S/N $\geqslant$ 4. The associated log-likelihood is

$$\mathcal{L}(\boldsymbol{p}) = -\sum_j \left\{ \frac{(\mathcal{V}_j - \hat{\mathcal{V}}_j)^2}{2\Sigma_j^2} - \log\left(\frac{\Sigma_j}{\sigma_j}\right) + \log\left[\Delta_j \operatorname{erf}\left(\frac{\Delta_j}{\sqrt{2}\,\Omega_j}\right) - \frac{2\Omega_j}{\sqrt{2\pi}} e^{-\Delta_j^2/2\Omega_j^2}\right]\right\},$$
$$\text{where } \Sigma_j^2 = \sigma_j^2 \frac{1+\mathcal{V}_j^2}{1+\hat{\mathcal{V}}_j^2}, \quad \Delta_j = \frac{1+\mathcal{V}_j\hat{\mathcal{V}}_j}{1+\mathcal{V}_j^2},$$
$$\text{and } \Omega_j^2 = \frac{\sigma_j^2}{(1+\mathcal{V}_j^2)(1+\hat{\mathcal{V}}_j^2)}, \tag{18}$$

where the $\mathcal{V}_j$ and σ_j are the observed closure amplitudes, the $\hat{\mathcal{V}}_j(\boldsymbol{p})$ are the model visibility amplitudes given parameters $\boldsymbol{p}$ (with the functional dependence suppressed for clarity), and erf(x) is the error function. This differs from Equation (B17) by constant normalization factors. In the limit of $\sigma_j/\hat{\mathcal{V}}_j \to 0$, the third term vanishes. However, for $\mathcal{V}_j$ of order unity, Equation (18) does not reduce to a Gaussian distribution in any S/N limit.

Finally, we note that this approximation is significantly better when $\hat{\mathcal{V}}_j$ is small. Generally, the closure amplitudes can be constructed such that $\mathcal{V}_j < 1$, approximating this requirement. Currently, the user is expected to independently define the set of closure amplitudes such that this is true prior to reading them into THEMIS.

6.5. Interferometric Polarization Fractions

The interferometric polarization fraction provides a second example of a non-Gaussian likelihood available in THEMIS. As with the closure amplitude, the source of the non-Gaussianity is the presence of the ratio in their definition. This leads to an asymmetric likelihood with an extended tail toward large $\breve{m}$ that is also well approximated by a Gaussian quotient distribution for S/N $\geqslant$ 2, given in Equation (B10). That $\breve{m}$ is defined by the ratio of visibilities constructed simultaneously on the same baseline places an additional constraint on the likelihood, permitting it to accurately be described by a single noise parameter (Appendix B): for S/N $\geqslant$ 2, the Gaussian quotient distribution is accurate at all $\breve{m}$ to 13% for S/N = 2 and 6% for S/N $\geqslant$ 4. The associated log-likelihood is identical in form to Equation (18):

$$\mathcal{L}(\boldsymbol{p}) = -\sum_j \left\{ \frac{(\breve{m}_j - \hat{\breve{m}}_j)^2}{2\Sigma_j^2} - \log\left(\frac{\Sigma_j}{\sigma_j}\right) + \log\left[\Delta_j \operatorname{erf}\left(\frac{\Delta_j}{\sqrt{2}\,\Omega_j}\right) - \frac{2\Omega_j}{\sqrt{2\pi}} e^{-\Delta_j^2/2\Omega_j^2}\right]\right\},$$
$$\text{where } \Sigma_j^2 = \sigma_j^2 \frac{1+\breve{m}_j^2}{1+\hat{\breve{m}}_j^2}, \quad \Delta_j = \frac{1+\breve{m}_j\hat{\breve{m}}_j}{1+\breve{m}_j^2},$$
$$\text{and } \Omega_j^2 = \frac{\sigma_j^2}{(1+\breve{m}_j^2)(1+\hat{\breve{m}}_j^2)}, \tag{19}$$

were $\breve{m}_j$ and σ_j are the observed polarization fraction and its uncertainty, and $\hat{\breve{m}}_j(\boldsymbol{p})$ are the model polarization fractions

[99] In Appendix B, we note that the appropriate log-normal distribution is a significantly worse approximation and biases image features reconstructed from data with moderate S/N values (4 < S/N $\lesssim$ 18).

associated with parameters $\boldsymbol{p}$ (with the functional dependence suppressed for clarity). This differs from Equation (B10) by constant normalization factors. As with the closure amplitudes, this is non-Gaussian even in the limit of $\sigma_j/\breve{m}_j \to 0$.

6.6. Norm-marginalized Visibility Amplitudes

Variations in the total source flux can be directly incorporated into the likelihood. Assuming Gaussian errors for visibility amplitudes, it is possible to introduce and analytically marginalize over an over-all normalization, V_{00}, presuming a flat prior (Broderick et al. 2014). This provides the maximum log-likelihood:

$$\mathcal{L}_{\max} = -\frac{\left(\sum_j V_j^2/\sigma_j^2\right)\left(\sum_j \hat{V}_j^2/\sigma_j^2\right) - \left(\sum_j V_j \hat{V}_j/\sigma_j^2\right)^2}{2\sum_j \hat{V}_j^2/\sigma_j^2}, \quad (20)$$

which may be identified with the minimum χ^2, occurring at

$$V_{00,\max} = \frac{\sum_j V_j \hat{V}_j/\sigma_j^2}{\sum_j \hat{V}_j^2/\sigma_j^2}. \quad (21)$$

More relevant for sampling is the marginalized log-likelihood:

$$\bar{\mathcal{L}} = \mathcal{L}_{\max} + \frac{1}{2}\log\left(\frac{2\pi V_{00,\max}^2}{\sum_j \hat{V}_j^2/\sigma_j^2}\right) \quad (22)$$

with the corresponding marginalized normalization, $V_{00,\mathrm{marg}} = V_{00,\max}$. A similar procedure could be applied for a Gaussian prior (see the following section) with very minor modifications.

By breaking the visibility amplitude data into epochs with similar visibility normalizations, corresponding, e.g., to a variable accretion rate, this can substantially increase the efficiency of sampling the remaining parameter space.

6.7. Shift-marginalized Closure Phases

At lowest order, refractive scattering induces shifts in the closure phase. These, again, may be incorporated into an appropriately constructed likelihood. Assuming Gaussian errors for closure phases, it is possible to analytically marginalize over an over-all shift, ϕ, presuming a user-supplied Gaussian prior (Broderick et al. 2016). This provides both the maximum log-likelihood,

$$\mathcal{L}_{\max} = -\frac{\phi_{\max}^2}{2\Sigma^2} - \sum_j \frac{\Delta^2(\Phi_j - \hat{\Phi}_j)}{2\sigma_j^2}, \quad (23)$$

where the most likely phase offset is

$$\phi_{\max} = \Sigma^2 \sum_j \frac{\Delta(\Phi_j - \hat{\Phi}_j)}{\sigma_j^2} \text{ where } \frac{1}{\Sigma^2} \equiv \sum_j \frac{1}{\sigma_j^2}, \quad (24)$$

and again $\Delta(x)$ is taken on $[-180°, 180°)$, minimizing the magnitude of the angular difference. The maximum log-likelihood is trivially related to the χ^2 and is relevant for fit quality assessment. More relevant for parameter estimation is the marginalized likelihood, for which the log-likelihood is given by

$$\bar{\mathcal{L}} = \mathcal{L}_{\max} - \frac{\phi_{\max}^2}{2(\sigma_\Phi^2 + \Sigma^2)} + \log\left(\frac{\Sigma}{\sqrt{\sigma_\Phi^2 + \Sigma^2}}\right), \quad (25)$$

with an associated marginalized value of the closure-phase shift of

$$\bar{\phi} = \frac{\sigma_\Phi^2}{\sigma_\Phi^2 + \Sigma^2}\phi^M. \quad (26)$$

Here, σ_Φ is the width of the Gaussian prior on ϕ; it is indicative of the amplitude of the refraction or turbulence responsible for the interepoch closure-phase fluctuations. This marginalized log-likelihood is appropriate for sampling the remaining parameters. By breaking the closure-phase data into epochs with similar visibility normalizations, corresponding, e.g., to a variable accretion rate, this can substantially increase the efficiency of sampling the remaining parameter space.

6.8. Gain-marginalized Visibility Amplitude

Station gains present a dominant source of systematic error in the construction of visibility amplitudes. Typically, it is possible to calibrate these to 10%–20% (Paper III); in instances where redundant antennas exist, network calibration schemes can reduce this to $\lesssim$1%. Occasionally, much larger gain variations are possible. In either case, these systematic errors overwhelmingly dominate the remaining, random components, e.g., the thermal noise. Accessing the full sensitivity of the EHT via visibility amplitude comparisons requires an efficient method of addressing the reconstruction and marginalization over these uncertain gains.

This reconstruction is facilitated by the correlations among visibilities resulting from the repeated appearance of the gain factors: in a scan involving N stations, at most N gains must be reconstructed along the underlying model from $N(N-1)/2$ visibility amplitude measurements. It is complicated by the time-variable nature of the gains, which can vary significantly from scan to scan, i.e., on timescales of tens of minutes, and thus introduces many additional parameters to be modeled (see, e.g., Natarajan et al. 2017). Within THEMIS, this is addressed by directly marginalizing over the gains at the level of likelihood construction. That is, THEMIS effectively self-calibrates the station gains to the model at each likelihood evaluation.

We assume that the gain at each station is well modeled by a set of constant corrections for each gain-reconstruction epoch. That is, for some set of $\{t_k\}$ for $t \in [t_k, t_{k+1})$, the gain-corrected model visibility amplitudes obtained from stations A and B are

$$|\hat{V}|_{AB,j}(g_{A,k}, g_{B,k}; \boldsymbol{p}) = (1 + g_{A,k})(1 + g_{B,k})|\bar{V}|(\boldsymbol{u}_{AB,j}; \boldsymbol{p}), \quad (27)$$

where $\bar{V}(\boldsymbol{u}; \boldsymbol{p})$ are the model amplitudes in the absence of gain corrections $g_{A,k}$ and $g_{B,k}$, and the baselines $\boldsymbol{u}_{AB,j}$ comprise those baselines connecting A to B for which measurements were made within the specified time frame. For each k, we reconstruct all of the gain corrections independently, subject to Gaussian priors on the $\boldsymbol{g}_k = (g_{A,k}, g_{B,k}, ...)$ with variance $\Sigma_{g,A}^2, \Sigma_{g,B}^2, ...$ This is done in two steps:

1. By numerically maximizing the log-likelihood in Equation (16) at fixed $\boldsymbol{p}$ supplemented by Gaussian priors for the coupled set of $\boldsymbol{g}_k$ to obtain $\bar{\boldsymbol{g}}_k$. This is necessarily peaked, unimodal, and a modest multidimensional optimization problem for the number of antennas relevant for the EHT. We do this using a modified Levenberg–Marquardt algorithm that includes the priors (see Appendix C).
2. An approximate marginalization over all $\boldsymbol{g}_k$ is performed via the Laplace approximation, i.e., by assuming that near $\bar{\boldsymbol{g}}_k$, the log-likelihood is nearly quadratic, and thus the contribution to the log-likelihood from $t \in [t_k, t_{k+1})$ is

$$\mathcal{L}_k(\boldsymbol{p}) \approx -\sum_j \frac{[|V|_j - |\hat{V}|_j(\bar{\boldsymbol{g}}_k; \boldsymbol{p})]^2}{2\sigma_j^2} - \frac{1}{2}\bar{\boldsymbol{g}}_k \cdot \boldsymbol{C} \cdot \bar{\boldsymbol{g}}_k - \frac{1}{2}\log||\boldsymbol{C}||, \tag{28}$$

where $\boldsymbol{C}$ is the covariance of the prior-adjusted log-likelihood from the first step and $||\boldsymbol{C}||$ is its determinant.

The final log-likelihood is obtained by combining those over all gain-reconstruction epochs: $\mathcal{L}(\boldsymbol{p}) = \sum_k \mathcal{L}_k(\boldsymbol{p})$.

The number of additional effective parameters introduced can be very large and depends on the particulars of the stations contributing during each gain-reconstruction epoch. Within each reconstruction epoch, the number of additional parameters is the minimum of the number of baselines and the number of stations participating. When all stations participate in N_g gain-reconstruction epochs, the number of additional model parameters is $N \times N_g$, which for a typical EHT observation in 2017 can be of order 10^2 for a single band. By ignoring potential correlations between gain-reconstruction epochs, we reduce the dimension of the additional parameter space by a large factor. By sampling the log-likelihood after marginalization over the $\{\boldsymbol{g}_k\}$, we efficiently restrict the dimension of the parameter space that must be sampled, e.g., by the techniques described in Section 7, to that of the original model. Typically, this has the effect of reducing the number of parameters by many hundreds.

We present validation tests in Section 8.2.2 and Paper VI.

7. Samplers

The dependence of the likelihood on the model parameters, incorporating any priors on the parameter values, is assessed via samplers. The process of sampling is conceptually separated from the definitions of data and models through the standardization of the likelihood objects. Thus, within THEMIS, a sampler is any method for exploring the values of a likelihood for various choices of the parameter vector. There is no standard output or input for a sampler, which may even vary qualitatively depending on the goal of the sampling process. However, all samplers interface with data and model objects solely through the use of likelihood objects and thereby permit analyses of a wide variety of combinations of data and models. Implemented samplers include the following.

7.1. Grid Search

The conceptually simplest but least efficient is a simple grid search where the parameter space is probed in predetermined fixed steps in each dimension. While limited in computational efficiency, this scheme is often used to cross-check results obtained by other samplers for smaller parameter spaces that both schemes can handle.

7.2. Parallel-tempered, Affine-invariant Markov Chain Monte Carlo

A natural choice for high-dimensional models is to use MCMC. Having scalability in mind, we chose to implement ensemble sampling methods in which many MCMC chains sample the parameter space in parallel. The chains interact and use the information from their spatial distribution to effectively adjust their next jump proposals. This has the added benefit of being able to sample the unknown likelihood surfaces efficiently and with minimal user input. We have implemented two different ensemble sampling methods, namely, an affine-invariant method and a differential-evolution method; the latter is given in Section 7.3.

The affine-invariant method can sample likelihood functions that are related by affine transformations with the same efficiency (Goodman & Weare 2010). This means it is very efficient in sampling highly stretched likelihood distributions as long as the nonlinear correlations among parameters are sufficiently weak.

MCMC algorithms are generally not very efficient on highly multimodal distributions. In order to overcome this problem, we have implemented parallel tempering for each MCMC sampler. Parallel tempering makes copies of the log-likelihood ($\mathcal{L}$) function that are made smoother through the introduction of a temperature parameter, the higher the temperature, the smoother the likelihood surface:

$$\mathcal{L}_i \propto \frac{\mathcal{L}}{T_i}. \tag{29}$$

The different temperatures are chosen from a temperature ladder such that $1 \leqslant T_i \leqslant T_{\max}$. Then, we run a copy of our MCMC sampler for each tempered likelihood copy in parallel. The highest temperature chains can freely move in the parameter space, while the low-temperature chains can be trapped in local likelihood maxima. By allowing the different temperature chains to exchange their positions following some prescription, the low-temperature chains efficiently escape the local maxima and explore the entire parameter space. The lowest-temperature chain, which samples the original untempered likelihood, yields the posterior probability distribution.

An efficient parallel-tempering algorithm requires that the temperature ladder has to be chosen carefully. There are two main factors to consider. First, the highest temperature used should be large enough to let the chains move freely within the likelihood surface. Furthermore, the temperatures should not be too widely spaced as that could hinder efficient swaps between chains from adjacent temperatures and lead to inefficient tempering. The choice of an efficient temperature ladder depends on the likelihood surface and could be difficult to guess. To mitigate this problem, we have implemented a method to adaptively change the temperatures in order to get near optimal efficiency. Our method follows that of Vousden et al. (2016).

The temperatures of the top and bottom tempering levels are fixed to $T = \infty$ and $T = 1$, respectively, ensuring that the top and bottom tempering levels sample the prior and posterior of interest. Intervening tempering levels are initialized to have a geometric ladder with the total number of levels, initial spacing

ratio, and tempering schedule treated as hyperparameters. During sampling runs, the user is responsible for inspecting the evolution of the resulting MCMC chains, tempering level exchange rates, and tempering level temperature evolution to ensure that the tempering is performing well. Key aspects to ensure are that the tempering level exchange rates are similar across the ladder (including to the prior), that the tempering level temperatures have settled to asymptotic values, and that the minimum number of tempering levels necessary are being used (thereby minimizing the roundtrip time). These must be explored via the comparison of multiple sampling runs, though they are frequently clearly discernible early in the run.

We have implemented parallelism in different levels of the MCMC sampler. Tempering is parallelized, MCMC chains at each tempering level run in parallel, and the likelihood can itself use multiple threads to run. This allows for the effective use of large high-performance computing machines.

It is a well-known feature of MCMC schemes that there is an initial so-called "burn-in" phase when the sampling exhibits comparatively large changes in parameter predictions, followed by a phase where the walkers settle down to smaller and more consistent changes, before ultimately converging to a final answer. As is customary in this approach, we exclude a certain number of MCMC steps corresponding to the "burn-in" process at the beginning of the "chains," i.e., the history of an MCMC walker. This "burn-in" is identified by inspection, although tools and diagnostics (e.g., Vehtari et al. 2019) are in active development and will be available in a future release of THEMIS.

7.3. Parallel-tempered, Differential-evolution Markov Chain Monte Carlo

The second MCMC algorithm implemented in THEMIS is the parallel-tempered, differential-evolution algorithm (Braak 2006). The differential-evolution method adjusts the collective move of its chains in a way to achieve optimal acceptance rate during Monte Carlo steps. It has the added benefit of being able to jump between modes in multimodal problems even without any tempering, thus representing a better option for multimodal distributions. Our implementation follows that of Nelson et al. (2014), and it also makes use of the same parallel-tempering algorithm as the affine-invariant method. Nelson et al. (2014) uses the ensemble to drive the acceptance rate of the sampler to a predetermined value, typically 0.25. Additionally, every 100 steps, it proposes a large step to potentially enable mode swapping for multimodel posteriors. Note, however, that this depends on the posterior having a regular structure.

7.4. Bayesian Evidence

The MCMC sampling described above provides the posterior probability distribution on parameters within the context of a given model. This allows us to calculate expectation values for any quantity of interest and to assess the goodness of fit for a given model to the data. However, if we need to compare the plausibility of different models given the same data set, running conventional MCMC is insufficient. Potential ways to perform model comparisons include reversible jump MCMC, calculating the Bayesian evidence (via thermodynamic integration, nested sampling, or Laplace approximation), and the use of information criteria (Knuth et al. 2015).

In Bayesian probability theory the relative probability of two models, M_1 and M_2, given the same data set, D, is related to the ratio of the Bayesian evidences for the two models, i.e., the Bayes factor or odds ratio. The relative posterior probability of the two models can be written as

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(D|M_1)}{P(D|M_2)}\frac{P(M_1)}{P(M_2)}, \tag{30}$$

$P(D|M_1)$ and $P(D|M_2)$ are the Bayesian evidence for the two models. In the absence of some strong prior preference, it is often assumed that the prior probability of the two models, $P(M_1)$ and $P(M_2)$, are equal, and hence the ratio of the Bayesian evidence or the Bayes factor is all that must be computed.

THEMIS implements the thermodynamics integration method to calculate the Bayesian evidence (Lartillot & Philippe 2006). In order to do this many MCMC chains are run in parallel on tempered versions of the likelihood. The temperature ladder for this purpose is provided by the user. Given the posterior distributions and the values of the likelihood at these points, the Bayesian evidence (Z) is obtained from

$$\ln Z = \int_0^1 E_\beta(\mathcal{L})\, d\beta, \tag{31}$$

where $\mathcal{L}$ is the log-likelihood, $\beta = 1/T$, and $E_\beta(\mathcal{L})$ is the expectation value of the log-likelihood calculated using the posterior probability distribution of chains at a tempered level corresponding to $T = 1/\beta$.

It is critical that the tempering ladder be sufficiently large to make contact with the prior, effectively sampled at $T = \infty$ ($\beta = 0$). Moreover, for the sum of the tempering levels to provide an accurate approximation of the Bayesian evidence, the separation between tempering levels must be sufficiently small. Both of these restrict the application of the Bayesian evidence, and often we will still appeal to information criteria.

8. Validation Tests

We now turn to validating THEMIS. For this we focus on the sampling methods, for which the implementations are novel, and reproducing prior analyses of EHT observations of Sgr A*. The variety in algorithmic improvements present in THEMIS results in considerable speed-up and simplicity in implementation in comparison to previous work to which we compare—often, analyses that took many months (e.g., Broderick et al. 2011, 2014, 2016) are now executed in days. Moreover, all of these tests have been integrated into THEMIS, both as validation tools and tutorials for future users.

Note that the sampler parameters (e.g., number of walkers, number of MCMC steps, number of tempering levels, grid search resolution) differ among these tests. In some cases, these are determined approximately algorithmically (e.g., the number of walkers in ensemble samplers should exceed the number of parameters by a factor of 4 or more; Braak 2006; Goodman & Weare 2010) while others are assessed by inspection (e.g., length of MCMC chains and grid search resolution). In principle, for problems that exhibit strong nonlinear covariances between all parameters, the number of walkers may scale as the number of parameters squared. However, in practice, the covariances for most problems are dominantly diagonal, i.e., the off-diagonals are sparse, and thus necessitate far fewer walkers.

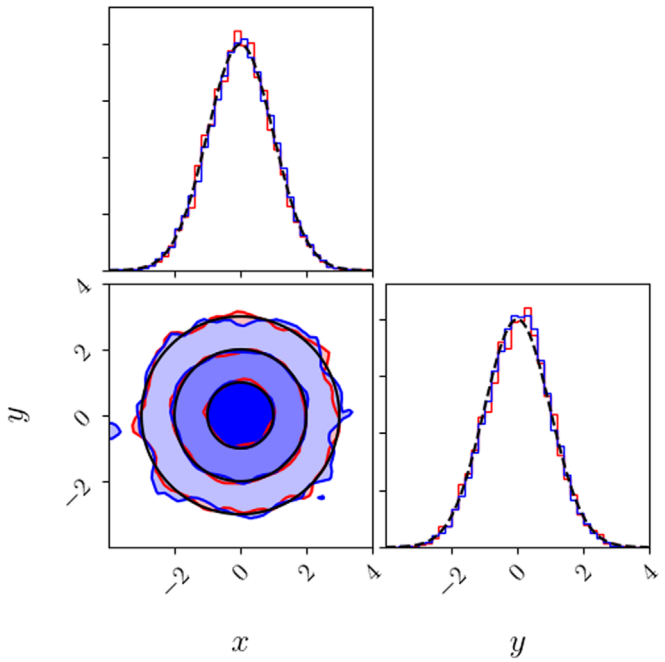

Figure 4. Two-dimensional Gaussian likelihood sampled by the affine-invariant MCMC sampler (blue), the differential-evolution MCMC sampler (red), and the grid search sampler (black).

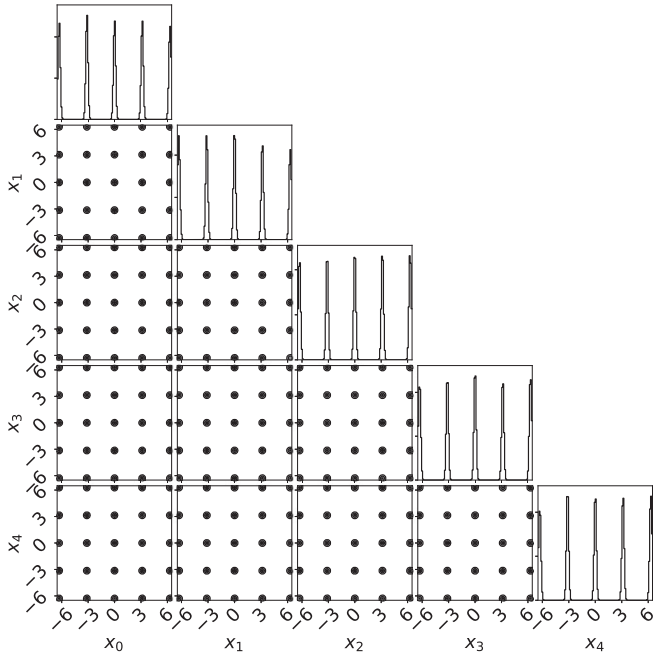

Figure 5. Validation of the sampler with a five-dimensional "egg-box"-likelihood test with the five artificial parameters x_0, x_1, x_2, x_3, and x_4. All peaks in the likelihood surface are successfully recovered.

8.1. Validation of the Samplers

Here we test the sampling part of the code thoroughly. In particular, we demonstrate the ability of the affine sampler to reliably probe nontrivial parameter spaces. That is, we anticipate that the complex models ultimately of most interest in the context of EHT analyses will produce multimodal probability distributions in high-dimensional parameter spaces. It will be necessary, therefore, to consistently identify all of the high-likelihood islands and determine accurately their relative posterior probabilities.

8.1.1. Two-dimensional Gaussian Likelihood

If our model has a small number of parameters, the grid search sampler can be efficient in sampling the parameter space. In this test, the grid search sampler was used to sample a two-dimensional symmetric Gaussian likelihood. Figure 4 shows the log-likelihood recovered using the grid search sampler as well as the marginalized posterior distributions for the same likelihood sampled using MCMC methods.

8.1.2. Egg Box Test

In this test, a five-dimensional parameter space with a highly multimodal egg-box-like distribution is sampled. The likelihood is described in Section 6.1 and contains $5^5 = 3125$ extrema, of which 1563 are maxima at the tops of sharp peaks within the prior range: $p_i \in [-8, 8]$ for all i.

This presents a significant challenge to most sampling schemes. The narrowness of the peaks and the dimensionality of the parameter space preclude a grid search, which would require more than 3×10^{12} samples to robustly detect all of them. The large dynamic range in the likelihood, i.e., the very low likelihoods between peaks, precludes typical MCMC schemes, which are unable to efficiently explore the full parameter space. Therefore, it provides a strong test of the ability of the parallel -tempered, affine-invariant, and differential-evolution parallel MCMC samplers to efficiently find and reconstruct the various high-likelihood regions.

As seen in Figure 5, both capture the gross features of the egg box likelihood. This run was executed employing the differential-evolution sampler with 8 tempering levels and took only 10 minutes on a typical laptop. There were 128 walkers running for 10^5 steps, of which the first 10^4 were discarded as the burn-in period

Every one of the 1563 maxima are populated, with an apparent average number of independent samples of only 10. This is sufficient to faithfully map out the diagonal elements of the covariance, which are reconstructed to better than 8% of the true value; the off-diagonal elements are restricted to less than 6% of the magnitude of the diagonal elements. While this is likely assisted by the highly symmetric nature of the peaks and their regular locations, it nevertheless demonstrates the ability of the samplers to effectively address highly multimodal likelihoods.

8.1.3. 16 Gaussian Test

Here we show that the sampling scheme can correctly probe a two-dimensional parameter space with a likelihood consisting of 16 Gaussians and accurately reconstruct the relative posterior probabilities of different peaks. To do this, the likelihood of the Gaussian located at $(x_0, x_1) = (20, 10)$ is chosen to be nine times higher than the others. This test was run using the affine-invariant sampler with four tempering levels. The number of MCMC steps in this case was 4000 steps, and the sampler used 100 walkers.

As shown in Figure 6, the sampler finds all of the Gaussian components. In addition, it recovers the nonuniformity of the

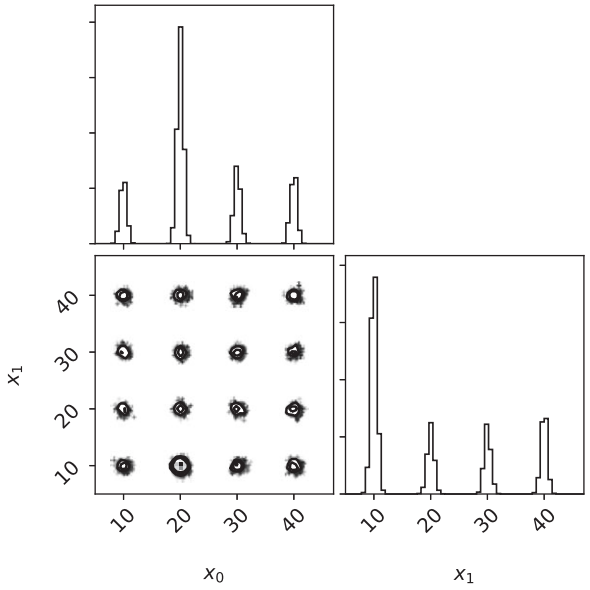

Figure 6. Validation of the sampler with a two-dimensional test involving a 16 Gaussian likelihood and 2 artificial parameters x_0 and x_1. Note that the sampler not only probes the likelihood surfaces comprehensively, it also correctly retrieves the Gaussian with the higher likelihood at $x_0 = 20$, $x_1 = 10$.

likelihood surface, accurately reconstructing the posterior weight of the appropriate component. Figure 7 shows the relative probability mass correctly recovered for these 16 Gaussian peaks.

8.1.4. 2D Rosenbrock Test

In this test, we demonstrate the differential-evolution sampler's ability to probe a two-dimensional parameter space with a likelihood surface that is quite narrow along one axis but broad in another and that has long tails. The two-dimensional Rosenbrock function serves as a standard test of sampling schemes. It has a global extremum at $(x_0, x_1) = (1, 1)$. We first ran the Rosenbrock test with flat priors and with three independent initial conditions for 100,000 steps with 24 walkers and temperatures. Upon inspection, these appear converged.

Given that sampling from the truncated Rosenbrock distribution is computationally much easier, we run again using a linear prior within the $[-2, 2]$ test and 20 walkers and 20 tempering levels for 5000 MCMC steps. As shown in Figure 8, the sampler probes the Rosenbrock likelihood surface well and recovers the global extremum. The fact that the marginalized distribution for x_1 peaks away from the truth is a result of the pathological functional form causing large contributions from $x_1 \sim 0$ regions. Note that the best fit maximizes the joint posterior distribution, not the marginalized distributions individually.

8.1.5. 5D Griewank Test

This test involves elements of both the egg box test and the Rosenbrock test in that it features many locally peaked likelihoods as in the egg box test and a long-tail component in the form of a parabola. The sampler finds the global extremum at $(x_i = 0)$ for $i = 0 \ldots 4$ and captures the parabolic falloff; see Figure 9. This test was performed by executing 3 independent runs using the differential-evolution sampler with 24 walkers and 24 tempering levels for 50,000 MCMC steps

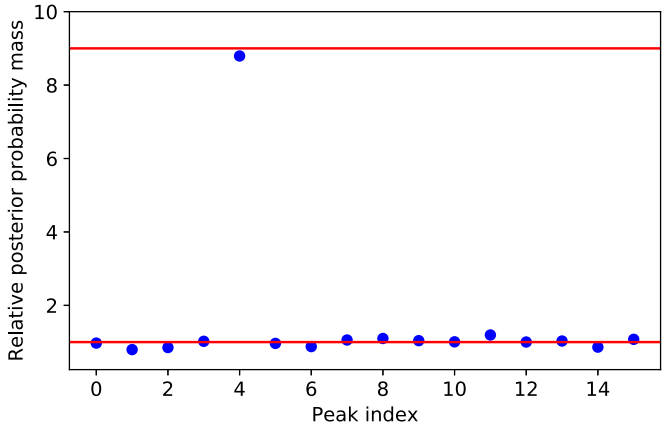

Figure 7. Relative posterior probability mass for the 16 Gaussian peaks recovered via MCMC sampling, in good agreement with the true values (red lines).

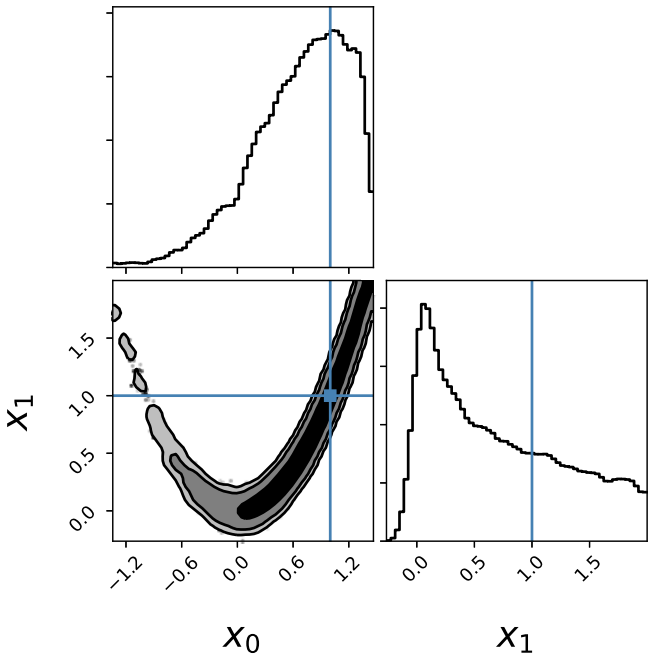

Figure 8. Validation of the differential-evolution sampler with a two-dimensional Rosenbrock test with two artificial parameters x_0 and x_1. The sampler finds the correct extremum and samples the likelihood surface comprehensively.

with flat priors. Upon inspection, the resulting chains appear well converged.

8.1.6. Multivariate Cauchy Distribution

The Cauchy (or Lorentz) distribution is a tough sampling problem due to its heavy tails and undefined mean. Its multivariate form, in particular, is therefore an effective and popular choice to test sampler performance. We ran the differential-evolution sampler on a three-dimensional Cauchy distribution using 24 walkers and temperatures for 100,000 MCMC steps with a flat prior, shown in Figure 10. While runs with a linear prior converge quickly, the flat prior runs require significantly longer evolutions. Even very long runs produce residual apparent structure in the resulting MCMC chains.

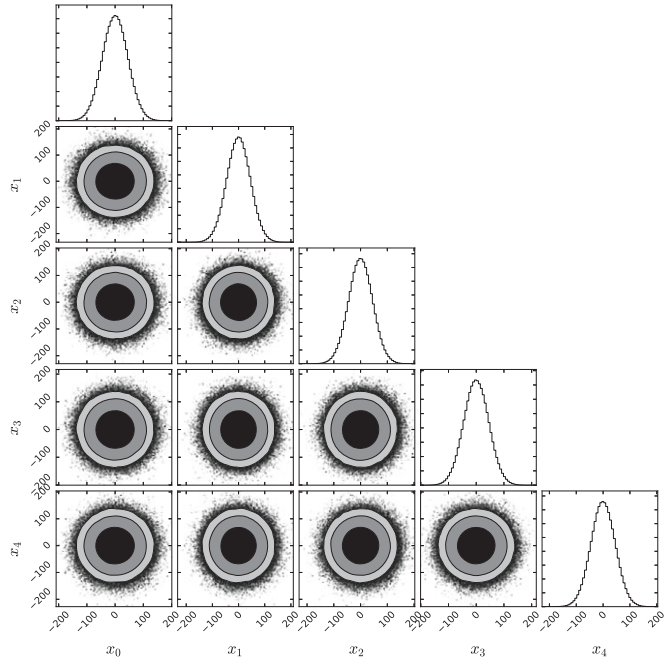

Figure 9. Validation of the differential-evolution sampler with a five-dimensional Griewank test with five artificial parameters x_0, x_1, x_2, x_3, and x_4. The sampler finds the correct extremum and explores the likelihood surface comprehensively.

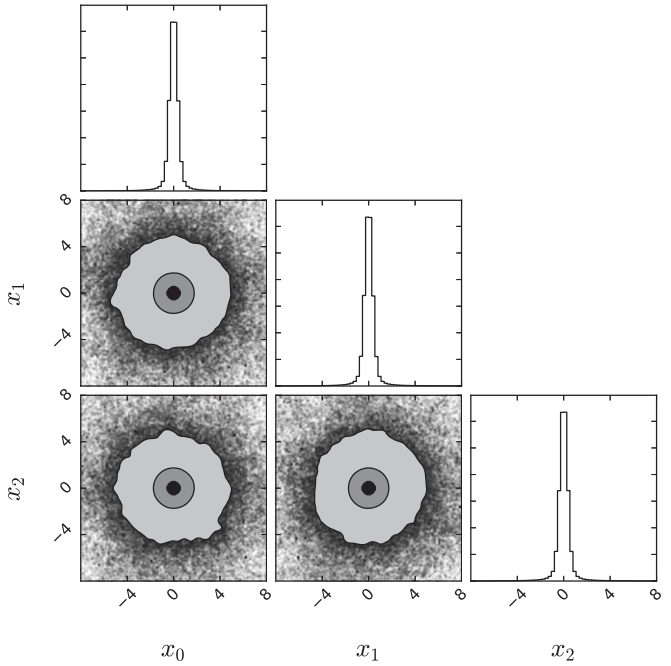

Figure 10. Validation of the differential-evolution sampler with a three-dimensional Cauchy/Lorentz test with three artificial parameters x_0, x_1, and x_2. The sampler finds the correct extremum and explores the likelihood surface comprehensively.

However, again, running an even slightly truncated version using linear priors (more akin to our typical challenges), convergence is achieved substantially more rapidly.

8.2. Self-tests with Simulated Data

Here, we demonstrate the ability of THEMIS to accurately reconstruct model parameters. This presents a simultaneous test of many of the components of THEMIS, including the data structures, models, likelihoods, and samplers. We generate simulated images using THEMIS' native model classes, from which the appropriate simulated data is constructed. Thermal noise is then included, producing a data set similar in character to that associated with a single night of the 2017 EHT campaign. The simulated data are then analyzed with THEMIS using the corresponding model. Note that this does not address model discrimination. For this purpose, we considered three models: the symmetric Gaussian and the geometric crescent models, for which visibilities can be computed analytically, and the SED-fitted RIAF model, which incorporates the ray-tracing components of VRT[2] and numerical data generation of THEMIS.

8.2.1. Gaussian Model

We generated simulated closure amplitude data from a compact, symmetric Gaussian with $V_0 = 2.5$ Jy and $\sigma = 5$ μas. We adopted a Gaussian for simplicity. The very compact size was selected to ensure high-S/N detections on even the longest baselines of the 2017 EHT campaign; such high S/Ns are typical of more complex models. We analyze these data with THEMIS' symmetric Gaussian model (Section 4.2.1) to assess potential biases associated with the non-Gaussian nature of the closure amplitude error distribution. We imposed an S/N minimum of 4 on the simulated closure amplitude.

For this analysis, we used the closure amplitude likelihood described in Section 6.4. We sampled the posterior distribution with the parallel-tempered, affine-invariant MCMC sampler, adopting linear priors on each model parameter. The analysis converged using 5 tempering levels with 128 walkers communicating every 50 MCMC steps and taking 5000 samples per walker.

As expected, the total intensity is not constrained by closure amplitudes, recovering our prior distribution. The resulting posterior distribution for the size of the Gaussian is shown in Figure 11. The reconstructed size is $\sigma = 5.0004 \pm 0.0004$ μas,[100] consistent with the input value. Repeating the analysis with different realizations of the simulated data produces qualitatively similar results, though they do exhibit 2σ fluctuations marginally more often than anticipated. No experiment produced a deviation larger than 3σ. As a result, we conclude that the likelihood in Equation (18) does not fully eliminate the bias inherent in the closure amplitude error distribution, though does so at the 2σ level. Decreasing the S/N minimum increases this bias substantially, suggesting that additional development is required to fully exploit low-S/N data.

While the closure amplitude likelihood in Equation (18) is not Gaussian, and thus does not admit a well-defined χ^2, we do construct an approximate expression via $\chi^2 = -2\mathcal{L}$. In the limit of small closure amplitudes, this identification is well motivated. The associated reduced χ^2 is 0.97 with 1656 degrees of freedom, suggesting that this statistic will be informative of fit quality. The corresponding two-tailed p-value, assuming a χ^2 distribution, is 39%.[101]

[100] Note that for all of our analysis, we report the marginalized values for the parameters of each model instead of the maximum values, which for complicated likelihood distributions may differ significantly.

[101] To quantitatively assess a given χ^2 value, we employ the two-tailed p-values, defined by $p = 2\min[P_<(\chi^2|\mathrm{dof}), P_>(\chi^2|\mathrm{dof})]$, where $P_{<,>}(\chi^2|\mathrm{dof})$ is the appropriate one-sided cumulative χ^2 for a given number of degrees of freedom, dof. This penalizes both poor fits ($\chi^2 \gg$ dof) and overfits ($\chi^2 \ll$ dof).

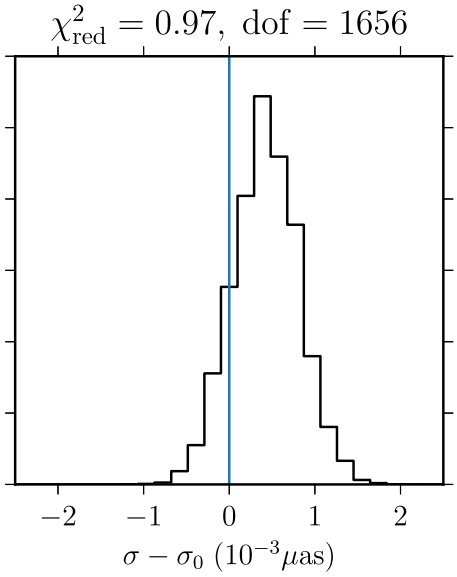

Figure 11. Posterior distribution of the size of the symmetric Gaussian reconstructed from simulated closure amplitude data with $\sigma_0 = 5\ \mu$as. The expected value is indicated by the vertical blue line.

8.2.2. Gaussian Model with Gain Errors

We generated simulated visibility amplitude data from a compact, symmetric Gaussian with $V_0 = 2$ Jy and $\sigma = 15\ \mu$as that incorporated known gain errors. The S/Ns, observing schedule, and baseline coverage approximate those associated with the 2017 April observations of M87.[102] Multiple types of gain errors were considered. Following the properties of the recent M87 observations, we permit potentially large errors on the LMT gains and modest errors on the remaining stations (Paper III). This presents a pathological situation, with one station considerably more poorly characterized. In all cases, the gains were assumed to be stationary across a scan.

The analysis employed the parallel-tempered, differential-evolution MCMC sampler, adopting linear priors on both. We adopt Gaussian priors on the reconstructed gain errors, with $\Sigma_g = 0.2$ for all stations with the exception of the LMT, for which $\Sigma_g = 1.0$. The analysis used 8 tempering levels, 128 walkers communicating every 50 MCMC steps, and converged rapidly.

The reconstructed gains for the best fit are shown in Figure 12 when the true gain errors vanish and when they are nonzero and vary throughout the night. In both cases, the same realization of the observational errors was employed. Modest fluctuations within the assumed gain-reconstruction priors driven by the thermal errors are present. This is clear from the analysis in which no gain errors were introduced. Patterns in the reconstructed gains persist across points within the MCMC chain and across different input gain-error realizations, indicating that they result from the overdetermined nature of the gain-reconstruction problem (see Section 6.8). Nevertheless, the reconstructions faithfully follow the introduced gain errors for all stations. This is most apparent for the LMT, which, by design, has the largest errors.

The process of reconstructing the gains does significantly expand the posteriors of the model parameters. This is clear in Figure 13, in which the red contours indicate the posterior

[102] Stations included are the SMA (SM), Pico Valeta (PV), LMT (LM), JCMT (JC), ARO-SMT (AZ), APEX (AP), and ALMA (AA).

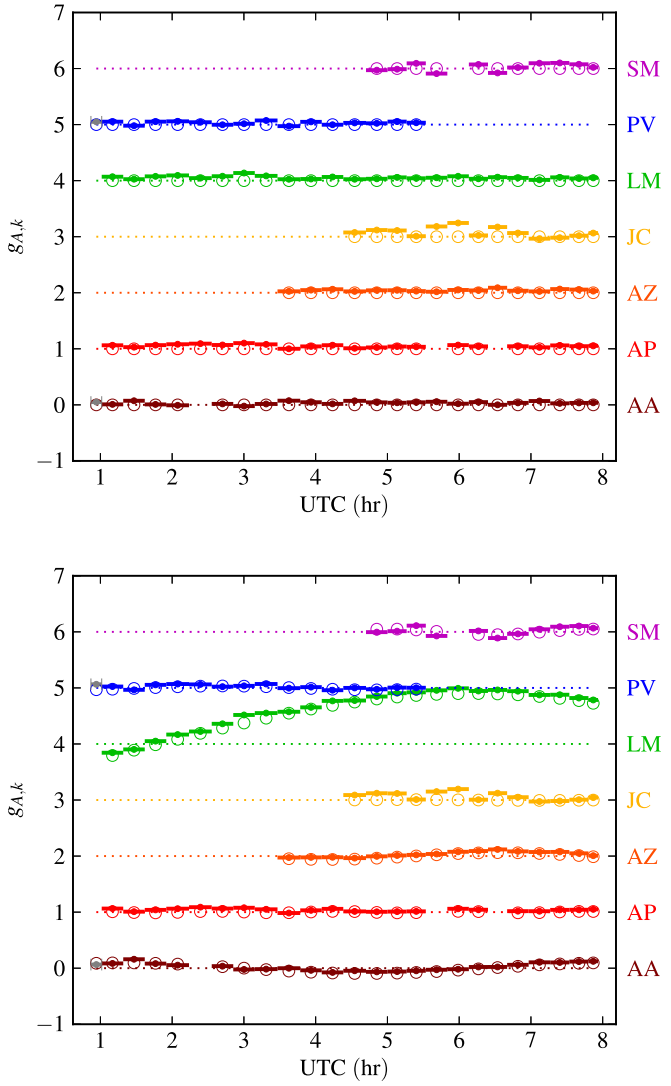

Figure 12. Reconstructed gains from simulated visibility amplitude data associated with a symmetric Gaussian with $\sigma_0 = 15\ \mu$as in the absence and presence of gain errors (top and bottom, respectively). Different stations are offset for clarity; for each station, the line of vanishing gain corrections is indicated by the horizontal dotted lines of the same color. Input gain corrections are shown by the open circles, and recovered gain corrections are shown by the filled dots with horizontal bars, which indicate the time range over which the gains are assumed to be fixed.

without gain reconstruction. This is expected: the additional freedom associated with the gain corrections can marginally correct for larger deviations from the true model. Nevertheless, the true model parameters are well within the posteriors generally. This is shown explicitly in Figure 13 for both cases shown in Figure 12.

8.2.3. Crescent Model

We generated simulated visibility amplitude and closure-phase data from a diffractively scattered crescent image with $V_0 = 2.24$ Jy, $R = 28\ \mu$as, $\psi = 0.14$, $\tau = 0.07$, and $\xi = 0°$, and added thermal noise to it. We then analyze these data with THEMIS' crescent model (Section 4.2.4), demonstrating that THEMIS properly recovers the parameters of the original image.

For the analysis, we used the standard visibility amplitude and closure-phase likelihoods described in Sections 6.2 and

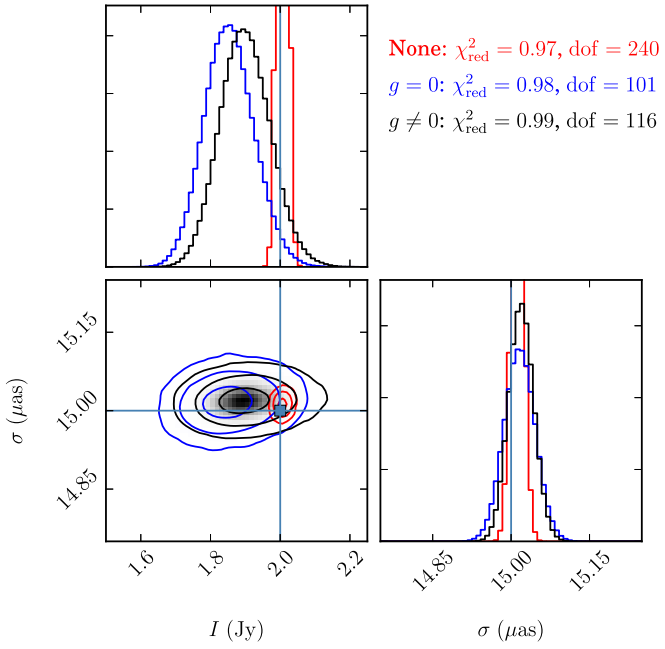

Figure 13. Posterior distribution of the size and amplitude of a symmetric Gaussian reconstructed from simulated visibility amplitude data with $\sigma_0 = 15$ μas without (red) and with (blue and black) station gain reconstruction. These include the posteriors from simulated data without gain errors (red and blue) and with significant imposed gain errors (black).

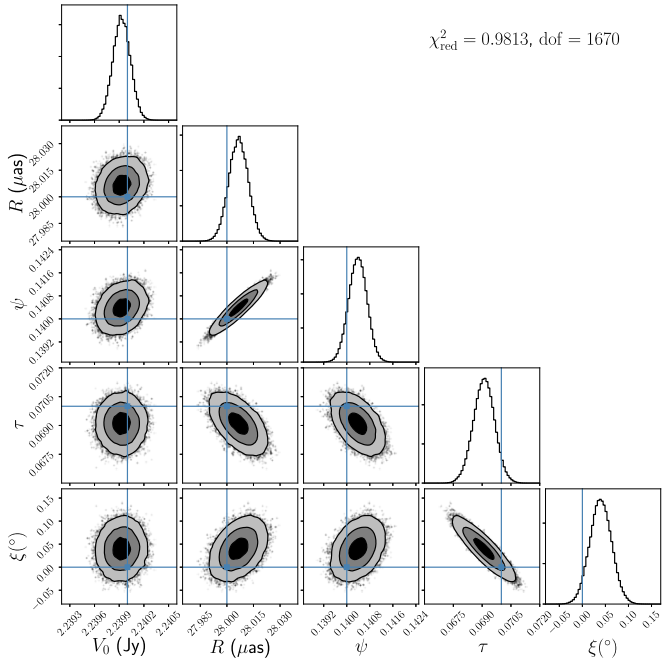

Figure 14. Joint parameter distributions from the crescent model analysis of the simulated visibility amplitude and closure-phase data generated from a crescent image with $V_0 = 2.24$ Jy, $R = 28$ μas, $\psi = 0.07$, $\tau = 0.14$, and $\xi = 0$ rad. Here, the true parameter values are represented by the blue lines, and the contours show to the 1σ, 2σ, and 3σ levels of the sampled posterior distribution.

6.3, and modeled the effects of diffractive scattering with the default scattering model implemented in THEMIS and described in Section 4.3.1.

In this case, we sampled the posterior distribution with the parallel-tempered, differential-evolution MCMC sampler adopting linear priors on each parameter of the model. The analysis converged using 4 tempering levels with 16 walkers per level communicating every 50 MCMC steps and taking 10,000 samples per walker. The resulting posterior distributions for the parameters of this model are shown in Figure 14 where the blue lines represent the true parameter values of the original image.

Our analysis shows that the marginalized values for the parameters of the model are $V_0 = 2.2399 \pm 0.0001$ Jy for the total flux, $R = 28.0064 \pm 0.0054$ μas for the overall size of the crescent, with a relative thickness $\psi = 0.1404 \pm 0.0003$, an asymmetry parameter $\tau = 0.0691 \pm 0.0005$, and a position angle $\xi = 0.^\circ040 \pm 0.^\circ023$. Individually, these are consistent at the 2σ level with the true values of the original crescent image. The model gives a satisfactory fit to the data as confirmed by the reduced χ^2 of 0.98 with 1670 degrees of freedom, which implies that high-quality fits exist. The corresponding two-tailed p-value is 59%, implying that the fit quality is well within the anticipated statistical range.

8.2.4. RIAF Model

We generated visibility amplitude and closure-phase data from a diffractively scattered RIAF image with (a, θ, ξ) = (0.10, 60°, 0°). We added thermal noise to the simulated data and then analyze it with THEMIS' SED-fitted RIAF model (Section 4.4.1) to show that THEMIS can properly recover the parameters of the original image.

For the analysis, the standard visibility amplitude and closure-phase likelihoods (Sections 6.2 and 6.3) were used, and the effects of diffractive scattering were modeled using THEMIS' default scattering model (Section 4.3.1). We used the parallel-tempered, differential-evolution MCMC sampler with 3 tempering levels, 14 walkers per level communicating every 50 MCMC steps, and took 5000 samples per walker.

The posterior distributions for the parameters of the model are shown in Figure 15. We find that the marginalized values for the black hole spin parameters are $a = 0.0997^{+0.0006}_{-0.0007}$, $\theta = 59.^\circ9983^{+0.^\circ0155}_{-0.^\circ0167}$, and $\xi = 0.^\circ0017^{+0.^\circ0223}_{-0.^\circ0199}$. These parameter estimates are consistent at the 1σ level with the true values of the original RIAF image. In this case, we find a reduced χ^2 of 0.99 with 1664 degrees of freedom and corresponding two-tailed p-value of 71%, indicating that high-quality fits were found.

8.3. Reproducing Previous Results

The variety of published analyses of EHT observations of Sgr A* provides a natural validation test of THEMIS, as well as a demonstration of its flexibility. These include comparisons of purely phenomenological and physically motivated models of the image structure. In constructing these, we make use of the published EHT data sets listed in bold in Table 1, consisting of visibility amplitudes measured in 2007 and 2009 and closure phases measured between 2009 and 2013, inclusively.

8.3.1. Symmetric Gaussian

We analyze the visibility amplitude data from 2007 and 2009 using the symmetric Gaussian model described in Section 4.2.1 in order to show that THEMIS can reproduce previous model-fitting studies made to estimate the source size of Sgr A*.

For the analysis, we employed the norm-marginalized visibility amplitude likelihood described in Section 6.6 to

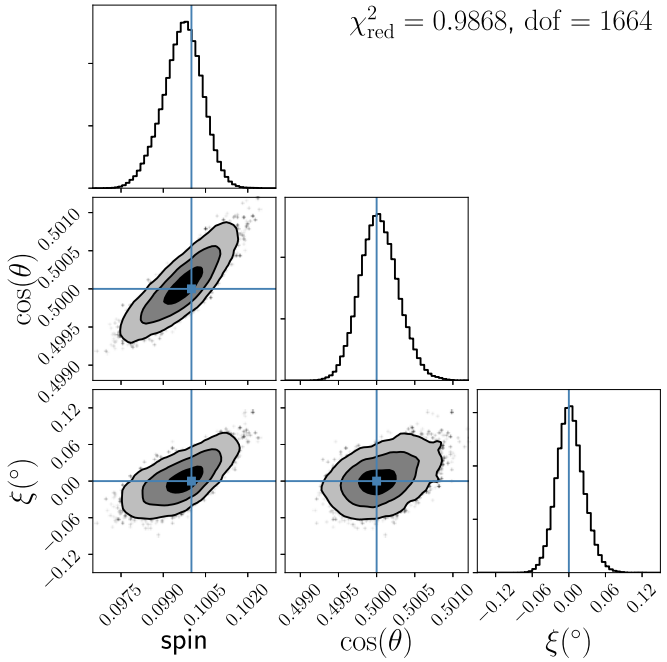

Figure 15. Joint parameter distributions for the SED-fitted RIAF model analysis of the simulated visibility amplitude and closure-phase data generated for a RIAF image with (a, $\cos\theta$, ξ) = (0.1, 0.5, 0.0). Here, the true parameter values are represented by the blue lines, and the contours show the 1σ, 2σ, and 3σ of the sampled posterior distribution.

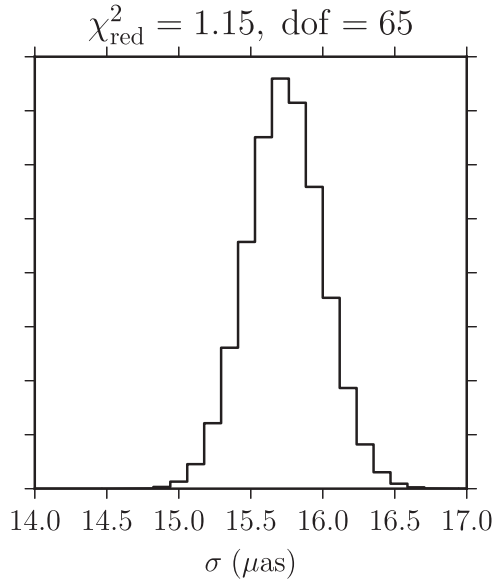

Figure 16. Gaussian size distribution from the analysis of the symmetric Gaussian model to the 2007 and 2009 visibility amplitude data of Sgr A*.

account for variations in the total flux of Sgr A* between observation nights. The effects of diffractive scattering were modeled using THEMIS' default scattering model (Section 4.3.1).

We employed the parallel-tempered affine-invariant sampler with 4 tempering levels with 32 walkers per level, and adopted linear priors on each parameter of this model. The posterior distribution for the intrinsic size of Sgr A* after 10,000 MCMC iterations is shown in Figure 16. The reconstructed size is $\sigma = 15.73 \pm 0.25$ (FWHM $= 37.05 \pm 0.60$). The model gives a satisfactory fit to the data with an associated reduced $\chi^2 = 1.15$ with 65 degrees of freedom, with a corresponding two-tailed p-value of 38%. These results are in good agreement with the best fits found in Broderick et al. (2011) when all epochs are combined and the inferred sizes for each night reported in Doeleman et al. (2008) and Fish et al. (2011).

8.3.2. Asymmetric Gaussian

We also analyze the visibility amplitude data from 2007 and 2009 with the asymmetric Gaussian model described in Section 4.2.2 to show that THEMIS can reproduce previous studies made to probe the asymmetry of the emitting region of Sgr A*.

For this analysis, we also analytically marginalized variations in the total flux of Sgr A* between observation nights using the norm-marginalized visibility amplitude likelihood described in Section 6.6, and modeled the effects of diffractive scattering using THEMIS' default scattering model (Section 4.3.1).

We employed the affine-invariant sampler and adopted linear priors on each parameter of this model. The results converge using 4 tempering levels, with 32 walkers per level, and taking 20,000 samples per walker. The posterior distributions for the different parameters of the model are shown in Figure 17. Our analysis shows that the marginalized values for the parameters of the model are $\sigma = 19.07^{+1.07}_{-2.51}$ μas, $A = 0.54^{+0.13}_{-0.18}$, and $\xi = -64.^\circ3^{+17.^\circ0}_{-4.^\circ7}$, and $\xi = 114.^\circ7^{+19.^\circ2}_{-5.^\circ1}$.

This model also gives a satisfactory fit to the data with an associated reduced $\chi^2 = 0.75$ with 63 degrees of freedom, with a corresponding two-tailed p-value of 14%. These results are in good agreement with the best fits found in Broderick et al. (2011) when all epochs are combined.

8.3.3. Crescent Model

We analyze the visibility amplitude data from 2007 and 2009 with the crescent model outlined in Section 4.2.4 in order to show that THEMIS can reproduce the earlier findings reported by Kamruddin & Dexter (2013).

We proceeded in a fashion similar to the analysis performed with the Gaussian models employing the norm-marginalized visibility amplitude likelihood described in Section 6.6 to account for variations in the total flux of Sgr A* between days and modeling the effects of diffractive scattering with the default scattering model implemented in THEMIS (Section 4.3.1).

In this case, we sampled the posterior distributions with the differential-evolution MCMC sampler adopting linear priors on each parameter of the model. We used the reported values in Table 1 of Kamruddin & Dexter (2013) as initial guesses for the values of the parameters of this model. The analysis converged using 6 tempering levels with 32 walkers per level, and taking 100,000 samples per walker. The resulting posterior distributions for the parameters of this model are shown in Figure 18. The marginalized values for the parameters of the model are $R = 29.8^{+4.8}_{-3.2}$ μas, $\psi = 0.28^{+0.12}_{-0.17}$, and $\tau = 0.20^{+0.15}_{-0.13}$. In this case, the analysis finds two values for the position angle $\xi = 60.^\circ5^{+8.^\circ4}_{-12.^\circ4}$, $\xi = -119.^\circ4^{+8.^\circ7}_{-13.^\circ7}$, and a minimum reduced $\chi^2 = 0.76$ with 62 degrees of freedom, with a corresponding two-tailed p-value of 16%.

The recovered crescent parameters are consistent with those reported in Kamruddin & Dexter (2013), though the inferred 1σ errors are larger. In addition, we recover the expected bimodal distribution in position angle, resulting from the insensitivity of

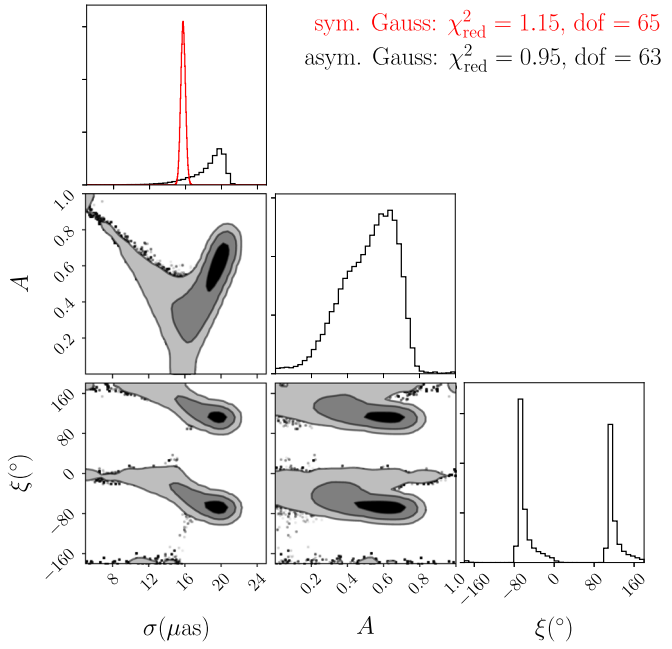

Figure 17. Joint parameter distributions from the analysis of the asymmetric Gaussian model of the 2007 and 2009 visibility amplitude data of Sgr A*. The gray contours show the 1σ, 2σ, and 3σ confidence regions for the size; the asymmetry parameter; and the position angle. For reference, the symmetric Gaussian size distribution from Figure 16 is shown in red in the top-left panel. These parameter distributions are consistent with the results of Broderick et al. (2011).

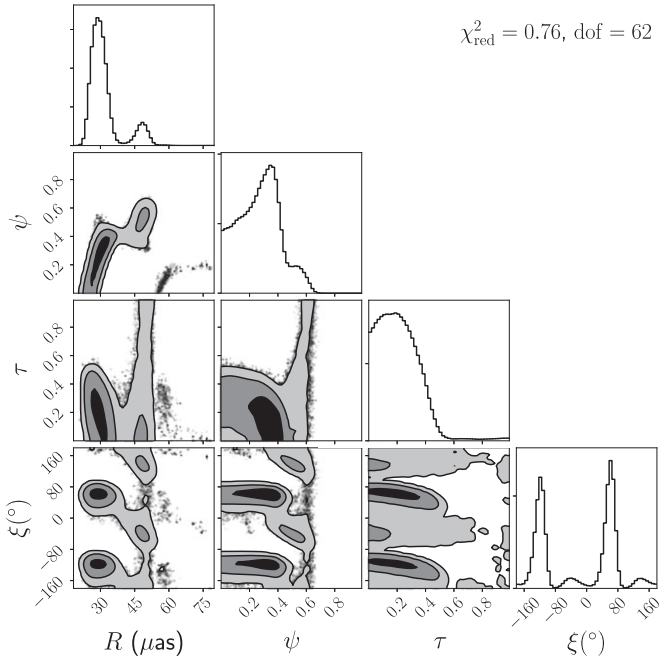

Figure 18. Joint parameter distributions from the analysis of the crescent model of the 2007 and 2009 visibility amplitude data of Sgr A*. The gray contours show the 1σ, 2σ, and 3σ confidence regions for the overall radius; the relative thickness; the degree of symmetry; and the position angle of the crescent. These parameter distributions are consistent with the results of Kamruddin & Dexter (2013).

the visibility amplitudes to a 180° rotation, and an additional weak bimodality in the crescent radius.

8.3.4. RIAF Model: Visibility Amplitude Analysis

We now turn to the first example of a physical model. First, we demonstrate THEMIS' ability to reproduce the analysis published in Broderick et al. (2011). For that purpose, we analyze the visibility amplitude data of Sgr A* from 2007 and 2009 with the SED-fitted RIAF model described in Section 4.4.1, using a tabulated set of accretion flow parameters obtained at different black hole spins and inclinations—and distributed with THEMIS—that reproduces the observed SED of Sgr A*.

For this analysis, we employed a set of linear priors for each parameter of the model and the norm-marginalized visibility amplitude likelihood described in Section 6.6 to account for variations in the total flux of Sgr A* between observation nights. The effects of diffractive scattering were modeled using THEMIS' default scattering model (Section 4.3.1).

We used the parallel-tempered, differential-evolution MCMC sampler with 12 tempering levels and 8 walkers per level communicating every 50 MCMC steps. The test completed 20,000 MCMC iterations, and the posterior distributions for the black hole spin parameters are shown in Figure 19. We find a spin $a = 0.25^{+0.35}_{-0.19}$, while the inclination angle has two values located at $\theta = 118\overset{\circ}{.}94^{+8\overset{\circ}{.}77}_{-10\overset{\circ}{.}02}$ and $\theta = 62\overset{\circ}{.}31^{+10\overset{\circ}{.}54}_{-9\overset{\circ}{.}33}$. The model has a minimum reduced $\chi^2 = 0.80$ with 63 degrees of freedom, with a corresponding two-tailed p-value of 25%. These results are in good agreement with the best fits found in Broderick et al. (2011) when all epochs are combined.

8.3.5. RIAF Model: Visibility Amplitude and Closure-phase Analysis

We analyze the visibility amplitude and closure-phase data sets that are bolded in Table 1 with the SED-fitted RIAF model described in Section 4.4.1 using 128 × 128 pixel RIAF images to show that THEMIS successfully reproduces the results of the analysis published by Broderick et al. (2016).

For this analysis, we employed the norm-marginalized visibility amplitude likelihood described in Section 6.6 to account for variations in the total flux of Sgr A* between observation nights. We also used the shift-marginalized closure-phase likelihood in Section 6.7 to model the effects of refractive scattering, while the effects of diffractive scattering were modeled using THEMIS' default scattering model (Section 4.3.1).

We used the parallel-tempered, differential-evolution MCMC sampler with 12 tempering levels and 8 walkers per level communicating every 50 MCMC steps. The MCMC chain was run for 20,000 steps, and the resulting posterior distributions for the parameters of this model are show in Figure 20 in comparison to the results of analysis with visibility amplitude data only discussed in the previous section. We find that the black hole spin parameters are similarly constrained after the inclusion of the closure-phase data, with $a = 0.09^{+0.11}_{-0.07}$, $\theta = 119\overset{\circ}{.}00^{+2\overset{\circ}{.}74}_{-2\overset{\circ}{.}12}$, and $\theta = 61\overset{\circ}{.}55^{+1\overset{\circ}{.}97}_{-3\overset{\circ}{.}25}$, and $\xi = -165\overset{\circ}{.}13^{+7\overset{\circ}{.}42}_{-4\overset{\circ}{.}62}$. In this case, we find a reduced χ^2 of 1.08 with 231 degrees of freedom, with a corresponding two-tailed p-value of 39%, indicating that high-quality fits were found. These results are consistent with best-fit parameters reported by Broderick et al. (2016).

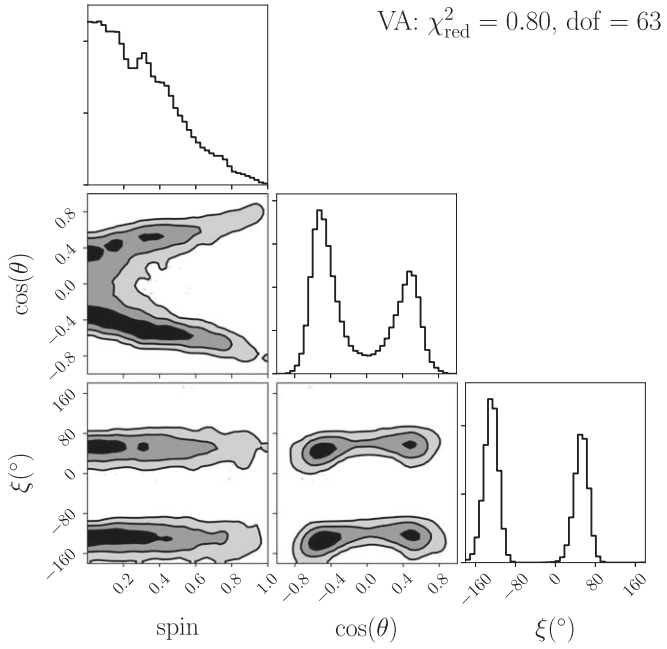

Figure 19. Joint parameter distributions from the analysis of the SED-fitted RIAF model of the 2007 and 2009 visibility amplitude data of Sgr A*. The gray contours show the 1σ, 2σ, and 3σ confidence regions for the spin magnitude, a, the cosine of the inclination, $\cos\theta$, and the position angle, ξ. These parameter constraints are consistent with the results of Broderick et al. (2011).

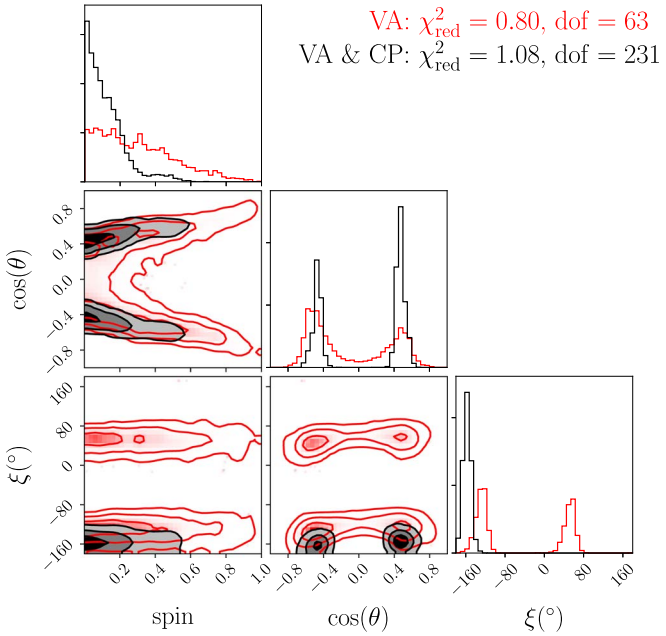

Figure 20. Joint posterior parameter distributions for the SED-fitted RIAF model implied by the combined visibility amplitude and closure-phase data sets bolded in Table 1. For reference, the posteriors implied by the visibility amplitude data alone are shown in red. The contours show the 1σ, 2σ, and 3σ confidence regions for the spin magnitude, a; the cosine of the inclination, $\cos\theta$; and the position angle, ξ. All parameter constraints are consistent with the results of Broderick et al. (2016).

9. New Results

In this section, we present novel results obtained with THEMIS. These include model comparisons to larger collections of legacy EHT data sets and/or more uniform comparisons than reported previously. Additional results obtained by combining a more complete combination of the data sets in Table 1 and applying additional model features will be reported elsewhere. Application to the 2017 EHT observations of M87 can be found in Papers V and VI.

9.1. Crescent Model

Prior analyses of the Kamruddin & Dexter (2013) crescent model utilized visibility amplitude data. With THEMIS, we extend these to include the closure-phase data sets that are bolded in Table 1. To account for refractive scattering, we employ the shift-marginalized closure-phase likelihood (Section 6.7) when including the contribution to the total likelihood from closure phases. In all other respects, the analysis is similar to that presented in Section 8.3.3.

The inclusion of closure-phase data places strong new constraints on the crescent structure in a number of respects. The resulting posterior distributions for the parameters of this model are shown in Figure 21. The constraints on all of the crescent parameters are substantially improved quantitatively, often settling ambiguities in the previous analysis. The overall crescent size has been restricted to $R = 46.3^{+1.4}_{-1.5}$ μas, the relative thickness parameter is now $\psi = 0.41^{+0.07}_{-0.04}$, and the asymmetry parameter is $\tau = 0.23^{+0.21}_{-0.14}$. Individually, these are consistent at the 2σ level with the expectation based on visibility amplitudes alone.

Similarly, the position angle is also strongly constrained, with the prior degeneracy eliminated, finding $\xi = 179.^\circ4^{+19.^\circ2}_{-9.^\circ1}$. Unlike the other parameters, this is inconsistent with the

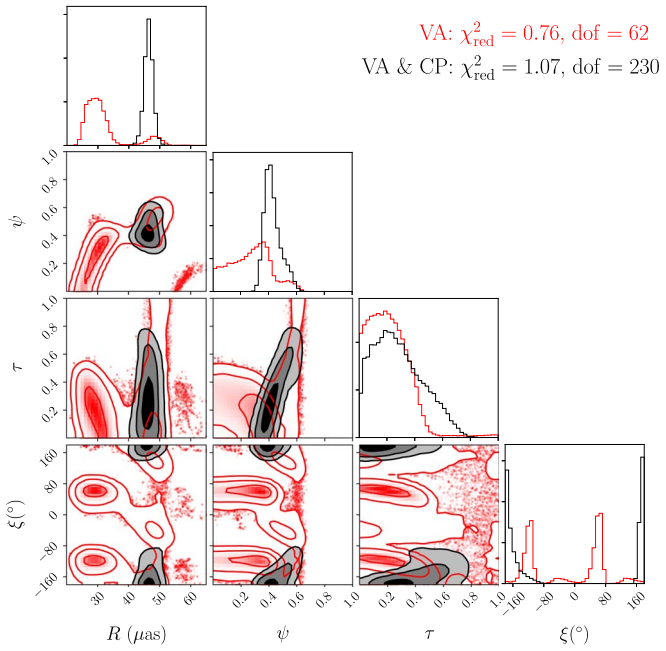

Figure 21. Joint posterior parameter distributions for the crescent model implied by the combined visibility amplitude and closure-phase data sets bolded in Table 1. For reference, the posterior distributions implied by the visibility amplitude data alone from Figure 18 are shown in red. The gray contours show the 1σ, 2σ, and 3σ confidence regions for the overall radius; the relative thickness; the degree of symmetry; and the position angle of the crescent.

estimates from the visibility amplitudes alone at the 2σ level. This is apparent in the bottom panels of Figure 21. This is modestly disconcerting, given the qualitatively distinct natures

of the closure phases and visibility amplitudes. Nevertheless, the reduced $\chi^2 = 1.07$, with a corresponding two-tailed p-value of 44%, implies that high-quality fits exist.

9.2. Extended RIAF Model

The SED-fitted RIAF model treats the comparisons to the EHT data and flux measurements differently, utilizing a prior set of compact radio SED fits. Here, we use THEMIS to relax this procedure, comparing RIAF models simultaneously to the flux and millimeter-VLBI measurements. In principle, this may broaden the black hole parameter estimates, trading worse compact radio SED fits for better structural fits. To explore this, we performed a new analysis, similar in spirit to that presented in Section 8.3.5, in which we analyze both data sets concurrently.

We performed a new analysis using the extended RIAF model described in Section 4.4.2, which generates flux measurements in addition to the millimeter-VLBI observations in a fashion identical to the SED-fitted RIAF model. In addition to the parameters describing the black hole spin (magnitude, inclination, and position angle), three additional parameters were introduced, describing the normalizations of the densities ($n_{e,t}$, $n_{e,\rm nt}$) and temperature (T_e) of the emitting electron population; all remaining parameters were held fixed at the values employed in the SED-fitted RIAF model: $\eta_t = -1.1$, $\eta_{\rm nt} = -2.02$, $\tau_t = -0.84$, $h_t = h_{\rm nt} = 1.0$, $\alpha = 1.25$, $\gamma_{\rm min} = 100$, $\beta = 10$, and $\kappa = 0$. This model was compared to the flux and millimeter-VLBI data bolded in Table 1. For this run, the differential-evolution sampler with 12 tempering levels was used. There were 24 walkers used, and the MCMC chain was run for 8200 steps. The obtained reduced $\chi^2 = 1.06$, with a corresponding two-tailed p-value of 50%, implying that high-quality fits were found.

In Figure 22, the resulting set of parameter constraints is presented in comparison to the prior analyses described in Sections 8.3.4 and 8.3.5. In all cases, the spacetime parameters are consistent with those found previously. Including the flux data produces a marginally stronger constraint on the black hole spin, $a = 0.1^{+0.19}_{-0.08}$, arising from the systematic decrease in the quality of the compact radio SED fits at higher spins (which was ignored in the prior analyses). Nevertheless, as anticipated, the inclination constraints are broadened, permitting $\theta = 62.2^{+5.3}_{-4.6}$ and $\theta = 117.2^{+5.5}_{-6.2}$.

10. Code Performance

As seen in many of the validation tests and example analyses presented in Sections 8 and 9, even for models with modest numbers of parameters, it is typical for the posterior probability distributions to be multimodal. As the models increase in sophistication, introducing additional physical freedoms and addressing various systematic uncertainties, this problem will be compounded by the need to explore high-dimensional parameters spaces. This is further complicated by the computational expense of numerically generating images of realistic astrophysical models. As a result, THEMIS has been designed to exploit the proliferation of modern HPC systems. Here we discuss the ways in which this has been, and may be, implemented, along with a description of THEMIS's scaling efficiency, demonstrating that it can run efficiently on very large machines.

THEMIS explicitly supports parallelization via MPI and implicitly via OpenMP and CUDA. MPI parallelization has been implemented at a number of levels, including the samplers, likelihood evaluation, and model generation, permitting users maximum flexibility in distributing the computational workload of an analysis.

Both the parallel-tempered, affine-invariant and differential-evolution MCMC sampling algorithms are designed to exploit parallelization in two levels. First, the use of parallel-tempering levels may be further parallelized by assigning separate tempering levels to different collections of processors. Second, the use of ensemble methods may be trivially parallelized among the individual walkers. Our implementation of the ensemble sampler evolves half of the walkers simultaneously while using the other, nonevolving half to determine the next proposed jump. Each walker in the "active" set can be evolved on a separate CPU. Upon completion, the "active" and "passive" sets swap, and the process is repeated. The result is a set of samplers that can immediately utilize $N_T N_W/2$ processors, where N_T is the number of tempering levels and N_W is the number of walkers, typically many times the number of parameters.

Image generation is an intrinsically parallelizable task. The VRT^2 library already natively supports MPI parallelization and vectorization via OpenMP. On modern Xeon-based systems, VRT^2 can efficiently use $N_L = 32$ cores to produce 128×128 pixel images before ancillary memory and communication costs become significant. Odyssey employs GPUs via CUDA and provides an example of mixed MPI/GPU support within THEMIS. The performance of mixed MPI/GPU computation depends mainly on the number and specifications of the GPU cards and is less sensitive to the number of MPI cores. Users implementing new THEMIS models are provided an MPI communicator and are only responsible for determining if and how parallelization should be implemented in their instance; they will be able to automatically exploit parallelization at the other levels.

Figure 23 shows the scaling of THEMIS on a representative sample problem with $N_T = 4$ tempering levels, $N_W = 16$ MCMC walkers, and $N_L = 32$ processors per likelihood evaluation. For this case, THEMIS scales with 94% efficient to 32, 88% efficient at 512 cores, and 84% at 1024 cores. Note that even modest increases in problem complexity involving larger images or higher-dimensional parameter spaces require a larger set of walkers and tempering levels, and allow more processors per likelihood evaluation. Thus, the scaling efficiency of THEMIS will improve with problem size. Already, THEMIS can run efficiently on several thousand cores.

11. Summary

THEMIS provides a new framework in which to develop and implement analyses of EHT observations. By focusing on the construction of interfaces, THEMIS enforces a modularity that facilitates rapid future development, ensuring flexibility and permitting extensibility. This flexibility is illustrated by the existing set of current THEMIS components, which span a wide variety of types of data, models, and sampling techniques. The clear definition of component inputs and outputs enables future developers to rapidly contribute additional components (e.g., image models) without the need for a global understanding of the internal structure of the code.

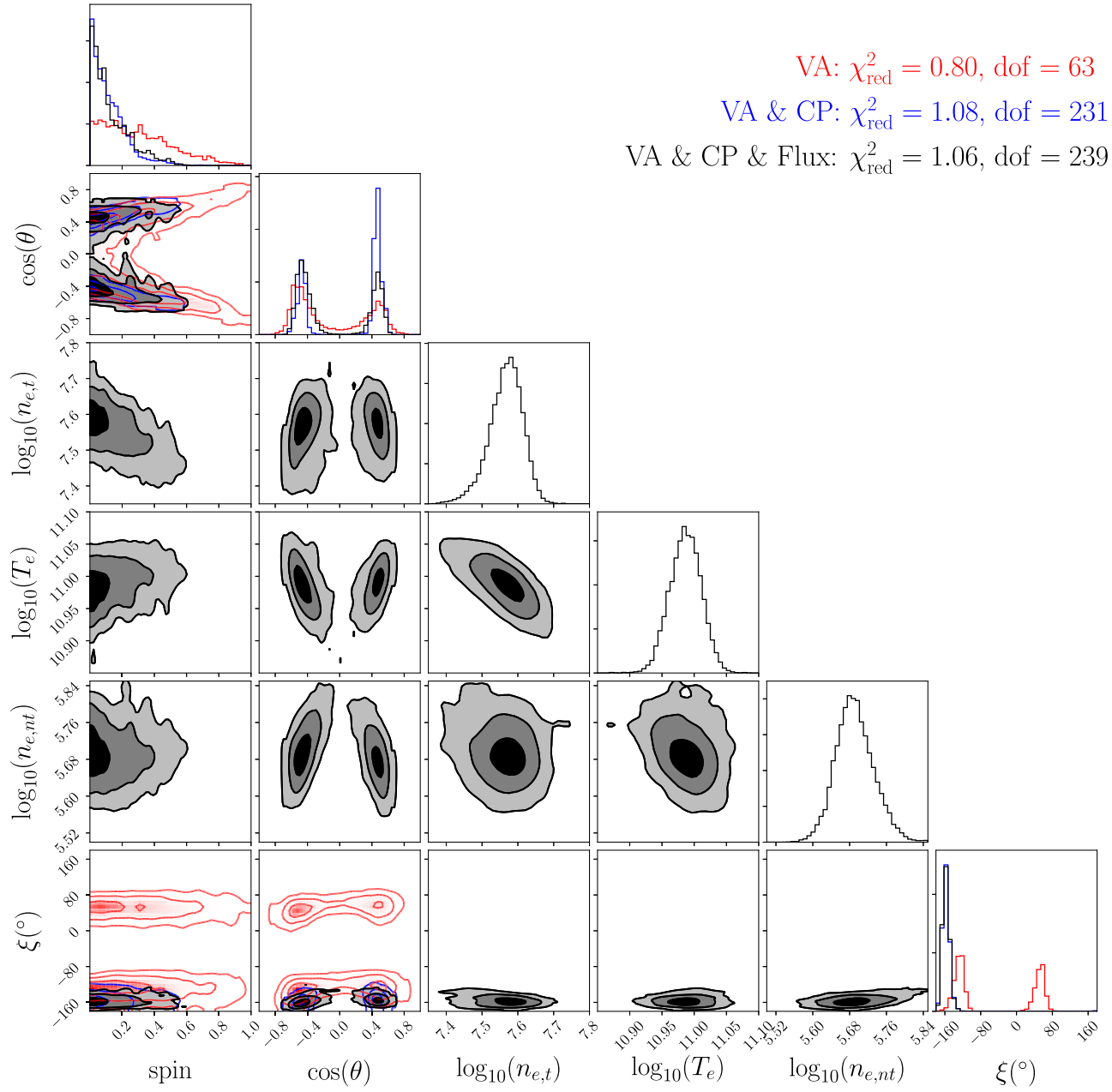

Figure 22. Joint posterior parameter distributions for the SED-fitted RIAF model implied by the combined flux, visibility amplitude, and closure-phase data sets bolded in Table 1. For reference, the posteriors implied by the visibility amplitude data alone and by the combined visibility amplitude and closure-phase data sets from Figures 19 and 20 are shown in red and blue, respectively. The contours show the 1σ, 2σ, and 3σ confidence regions for the spin magnitude, a; the cosine of the inclination, $\cos\theta$; and the position angle, ξ. All parameter constraints are consistent with the results of Broderick et al. (2016).

Implemented data types include both millimeter-VLBI observables (visibility amplitudes, closure phases, closure amplitudes, polarization fractions) and ancillary data (fluxes). The ability to easily add accoutrements to these data objects, e.g., time stamps, observing stations, atmospheric conditions, observation resolution, etc., significantly increases their flexibility and the potential sophistication of subsequent analyses.

The generic nature of the model interface produces a correspondingly broad array of acceptable models, solving a key difficulty with unifying prior EHT analyses. As a result, THEMIS can construct analyses of phenomenological (e.g., Gaussians) and physically motivated models (e.g., polarized images of synchrotron-emitting GRMHD simulations). It also naturally allows the inclusion of optional, additional, independent model features (e.g., interstellar scattering) in a uniform way. In principle, it can also facilitate in nonparametric modeling, e.g., image inversion, though this has yet to be implemented.

A number of likelihoods have been implemented, including likelihoods that analytically address nuisance parameters. This will become increasingly important as additional EHT systematics are considered, e.g., telescope gain corrections, refractive scattering, and intrinsic source variability. Similarly, a number of samplers have been implemented, including samplers that efficiently explore high-dimensional, multimodal likelihood surfaces.

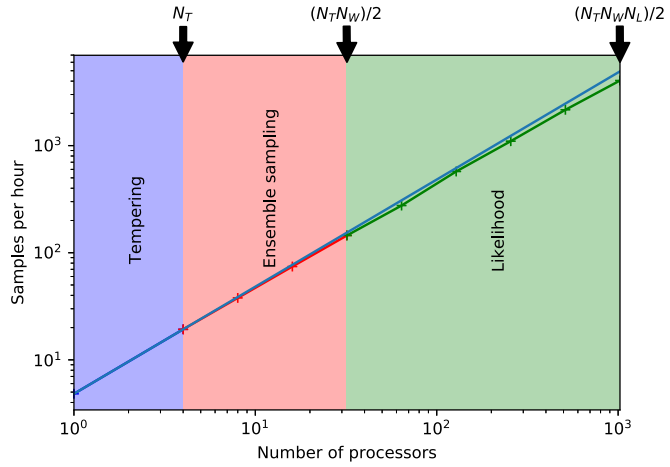

Figure 23. Scaling plot illustrating the multilayered parallelization strategy in THEMIS. The parallelization in tempering levels (perfect by construction) is shown in blue, the parallelization in different walkers is shown in red, and the parallelization in likelihood evaluations (e.g., images here via calling VRT[2]) is shown in green. The blue, red, and green data points show the scaling of the code. The blue line acts as a reference for perfect linear scaling.

A key feature of THEMIS is the ability to mix and match the above, constructing new analyses via minor changes in the model used, data included, and sampler used. This will be critical to evaluating the robustness of features, teasing apart subtle interactions in aspects of complex models, and systematically assessing the impact of additional types of data. At the same time, this permits rapid, distributed development: as features are added in the service of one analysis, e.g., a new sampler or a new scattering model, they may be rapidly deployed to others.

In anticipation of increasingly complex, physically motivated emission models for EHT targets, THEMIS enables the implementation of parallelization at multiple levels via multiple schemes. At present, this is implemented in a number of samplers via MPI and models via MPI, OpenMP, and CUDA. As a result, for typical analyses, THEMIS scales efficiently to thousands of cores, depending on problem complexity, and can effectively exploit modern HPC systems. For the implemented samplers, this parallel-performance scaling improves with problem complexity (i.e., number of parameters), partially mitigating the introduction of additional physical features.

Both the individual components of THEMIS and their integration have been extensively tested. THEMIS is able to accurately and consistently explore high-dimensional multimodal posterior probability distributions. It is able to recover the parameters of models used to construct realistic simulated EHT data for both geometric and physically motivated RIAF models. It has accurately reproduced previous analyses of published EHT data. In the case of the RIAF models, it has done so in an order-of-magnitude less user time.

The extensibility of THEMIS is evident in the extension of these prior analyses. The Kamruddin & Dexter (2013) crescent model has been reassessed in light of the EHT closure-phase measurements of Sgr A* published in Fish et al. (2016). The weak degeneracy in the size of the crescent is now broken, selecting $R = 46.3^{+1.4}_{-1.5}$ μas. This is considerably larger than the size implied by the 1σ region obtained when only visibility amplitude is considered, though still consistent at 2σ. Nevertheless, high-quality fits of the combined closure-phase and prior visibility amplitude data sets do exist. Note that this implies a crescent diameter that is nearly twice as large as the 55 μas anticipated for Sgr A* by identifying the crescent with the gravitationally lensed image of a geometrically thick accretion flow.

Where prior RIAF analyses have separated the fitting of the SED and EHT data for Sgr A*, THEMIS now simplifies the process of fitting both simultaneously. While this may yield weaker parameter constraints in principle, in practice the black hole spin parameters are similarly constrained. Within the context of a similarly constrained RIAF model, i.e., a geometrically thick, Keplerian flow, we find $a = 0.1^{+0.19}_{-0.08}$; $\theta = 62.2^{+5.3}_{-4.6}$, and $\theta = 117.2^{+5.5}_{-6.2}$; and $\xi = -158.^{\circ}1^{+11.^{\circ}5}_{-10.^{\circ}4}$. Future analyses that will systematically explore the relaxation of assumptions about the structure of the inner accretion flow will be published elsewhere.

THEMIS is meant to facilitate continuous, vigorous development. Already, plans are underway to address refractive scattering in the interstellar medium, model stochastic variability in the intrinsic emission region, introduce jet models, exploit GRMHD simulations, and perform nonparametric analyses. Anticipated future data-type development will include polarized fluxes, Faraday rotation measurements, circular polarization, and visibility variances. Additionally, new samplers will be implemented to help with the known difficulties of ensemble methods presented here (Huijser et al. 2015). As a result, THEMIS is designed to facilitate the future scientific utilization of the new window on black hole physics now opened by the EHT.

This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada (www.computecanada.ca). Computations were made on the supercomputer Mammouth Parallèle 2 from University of Sherbrooke, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), the ministère de l'Économie, de la science et de l'innovation du Québec (MESI), and the Fonds de recherche du Québec—Nature et technologies (FRQ-NT). This work was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. A.E.B. thanks the Delaney Family for their generous financial support via the Delaney Family John A. Wheeler Chair at Perimeter Institute. A.E.B., P.T., and M.K. receive additional financial support from the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant. R.G. receives additional support from the ERC synergy grant "Black-HoleCam: Imaging the Event Horizon of Black Holes" (grant No. 610058). We thank Jason Dexter for helpful comments and suggestions. We further thank the following organizations and programs: the Academy of Finland (projects 274477, 284495, 312496); the Advanced European Network of E-infrastructures for Astronomy with the SKA (AENEAS) project, supported by the European Commission Framework Programme Horizon 2020 Research and Innovation action under grant agreement 731016; the Alexander von Humboldt Stiftung; Black Hole Initiative at Harvard University, through grants from the John

Templeton Foundation and the Gordon and Betty Moore Foundation; the China Scholarship Council; Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile, via PIA ACT172033, Fondecyt 1171506, BASAL AFB-170002, ALMA-conicyt 31140007); Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico, projects 104497, 275201, 279006, 281692); the Delaney Family via the Delaney Family John A. Wheeler Chair at Perimeter Institute; Dirección General de Asuntos del Personal Académico–Universidad Nacional Autónoma de México (DGAPA–UNAM, project IN112417); the European Research Council Synergy Grant "BlackHoleCam: Imaging the Event Horizon of Black Holes" (grant 610058); the Generalitat Valenciana postdoctoral grant APOSTD/2018/177; the Gordon and Betty Moore Foundation (grants GBMF-3561, GBMF-5278); the Istituto Nazionale di Fisica Nucleare (INFN) sezione di Napoli, iniziative specifiche TEONGRAV; the International Max Planck Research School for Astronomy and Astrophysics at the Universities of Bonn and Cologne; the Jansky Fellowship program of the National Radio Astronomy Observatory (NRAO); the Japanese Government (Monbukagakusho: MEXT) Scholarship; the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for JSPS Research Fellowship (JP17J08829); the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS, grants QYZDJ-SSW-SLH057, QYZDJ-SSW-SYS008); the Leverhulme Trust Early Career Research Fellowship; the Max-Planck-Gesellschaft (MPG); the Max Planck Partner Group of the MPG and the CAS; the MEXT/JSPS KAKENHI (grants 18KK0090, JP18K13594, JP18K03656, JP18H03721, 18K03709, 18H01245, 25120007); the MIT International Science and Technology Initiatives (MISTI) Funds; the Ministry of Science and Technology (MOST) of Taiwan (105-2112-M-001-025-MY3, 106-2112-M-001-011, 106-2119-M-001-027, 107-2119-M-001-017, 107-2119-M-001-020, and 107-2119-M-110-005); the National Aeronautics and Space Administration (NASA, Fermi Guest Investigator grant 80NSSC17K0649); the National Institute of Natural Sciences (NINS) of Japan; the National Key Research and Development Program of China (grant 2016YFA0400704, 2016YFA0400702); the National Science Foundation (NSF, grants AST-0096454, AST-0352953, AST-0521233, AST-0705062, AST-0905844, AST-0922984, AST-1126433, AST-1140030, DGE-1144085, AST-1207704, AST-1207730, AST-1207752, MRI-1228509, OPP-1248097, AST-1310896, AST-1337663, AST-1440254, AST-1555365, AST-1715061, AST-1615796, AST-1716327, OISE-1743747, AST-1816420); the Natural Science Foundation of China (grants 11573051, 11633006, 11650110427, 10625314, 11721303, 11725312); the Natural Sciences and Engineering Research Council of Canada (NSERC, including a Discovery Grant and the NSERC Alexander Graham Bell Canada Graduate Scholarships-Doctoral Program); the National Youth Thousand Talents Program of China; the National Research Foundation of Korea (the Global PhD Fellowship Grant: grants NRF-2015H1A2A1033752, 2015-R1D1A1A01056807, the Korea Research Fellowship Program: NRF-2015H1D3A1066561); the Netherlands Organization for Scientific Research (NWO) VICI award (grant 639.043.513) and Spinoza Prize SPI 78-409; the New Scientific Frontiers with Precision Radio Interferometry Fellowship awarded by the South African Radio Astronomy Observatory (SARAO), which is a facility of the National Research Foundation (NRF), an agency of the Department of Science and Technology (DST) of South Africa; the Onsala Space Observatory (OSO) national infrastructure, for the provisioning of its facilities/observational support (OSO receives funding through the Swedish Research Council under grant 2017–00648) the Perimeter Institute for Theoretical Physics (research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research, Innovation and Science); the Russian Science Foundation (grant 17-12-01029); the Spanish Ministerio de Economía y Competitividad (grants PGC2018-098915-B-C21, AYA2016-80889-P); the State Agency for Research of the Spanish MCIU through the "Center of Excellence Severo Ochoa" award for the Instituto de Astrofísica de Andalucía (SEV-2017-0709); the Toray Science Foundation; the US Department of Energy (USDOE) through the Los Alamos National Laboratory (operated by Triad National Security, LLC, for the National Nuclear Security Administration of the USDOE (Contract 89233218CNA000001)); the Italian Ministero dell'Istruzione Università e Ricerca through the grant Progetti Premiali 2012-iALMA (CUP C52I13000140001); the European Union's Horizon 2020 research and innovation program under grant agreement No. 730562 RadioNet; ALMA North America Development Fund; the Academia Sinica; Chandra TM6-17006X. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by NSF grant ACI-1548562, and CyVerse, supported by NSF grants DBI-0735191, DBI-1265383, and DBI-1743442. XSEDE Stampede2 resource at TACC was allocated through TG-AST170024 and TG-AST080026N. XSEDE JetStream resource at PTI and TACC was allocated through AST170028. The simulations were performed in part on the SuperMUC cluster at the LRZ in Garching, on the LOEWE cluster in CSC in Frankfurt, and on the HazelHen cluster at the HLRS in Stuttgart.

Appendix A
Prediction Accuracy Requirements

When predictions are made numerically, frequently the computational expense is strongly dependent on the accuracy of the theoretical estimate required. Thus, significant efficiencies can be realized by understanding and limiting the accuracy requested where possible. Generally, comparisons with data with large uncertainties require far less accurate theoretical estimates than with data that have small uncertainties. Here we determine the relationship with parameter estimation uncertainty and the accuracy of the theoretical estimates, thereby estimating the accuracy required by THEMIS.

We begin by assuming that the measurement errors are Gaussian. We further assume that the posterior parameter probability is also nearly Gaussian and thus adopt a Fisher matrix approach to the estimation of the uncertainty of the parameter estimates. Finally, we assume that errors in the predicted values are Gaussian and uncorrelated. That is, we set the log-likelihood to

$$\mathcal{L} = -\sum_j \frac{(f_j + \delta_j - y_j)^2}{2\sigma_j^2}, \tag{A1}$$

where the predicted value is f_j, δ_j is the error in the predicted value, y_j are the data, and σ_j are the observational uncertainties.

The assumption that the prediction errors are Gaussian corresponds to assuming that the δ_j are Gaussian random variables. This is not true in an absolute sense: each time a prediction is made for the same independent variables, the δ_j

does not change. However, in a statistical sense, we are assuming that at different independent variable values and for different parameter values, the δ_j are well approximated by a random variable. It will be useful henceforth to characterize the size of the distribution of the δ_j in terms of σ_j, i.e., we set the variance of the prediction error in terms of the measurement uncertainty as $\langle \delta_j^2 \rangle = \Delta^2 \sigma_j^2$. This implies that higher prediction accuracy is possible for more accurately measured quantities.

The δ_j modify the minimum $\mathcal{L}$ (and thus χ^2) expected: averaging over realizations of the data and the prediction errors,

$$\langle \mathcal{L} \rangle = -\sum_j \left\langle \frac{(f_j - y_j)^2 + \delta_j^2}{2\sigma_j^2} \right\rangle \approx -\frac{N}{2}(1 + \Delta^2), \quad \text{(A2)}$$

where we have further assumed the number of degrees of freedom is close to the number of data points, N. For a sufficiently large number of degrees of freedom and a large enough Δ^2, the deviation will be statistically noticeable in the reduced χ^2 when $\Delta \gtrsim (8/N)^{1/4}$. This expression grows slowly with N, though, and is therefore not a fundamental limit as far as parameter estimation is concerned.

The uncertainty in the estimate of the parameters is set by the inverse of the covariance matrix, given by

$$C_{ab} = -\frac{\partial^2 \mathcal{L}}{\partial p_a \partial p_b} = \sum_j \frac{1}{\sigma_j^2} \times \left[\frac{\partial f_j}{\partial p_a} \frac{\partial f_j}{\partial p_b} + (f_j - y_j + \delta_j) \frac{\partial^2 f_j}{\partial p_a \partial p_b} \right]. \quad \text{(A3)}$$

Again, averaging over realizations of the data and the prediction errors gives

$$\langle C_{ab} \rangle = \sum_j \frac{1}{\sigma_j^2} \frac{\partial f_j}{\partial p_a} \frac{\partial f_j}{\partial p_b}. \quad \text{(A4)}$$

However, a typical value will be modified by the presence of the linear term. That is, the variance in the inverse covariance is

$$\langle (C_{ab})^2 \rangle - \langle C_{ab} \rangle^2 = \sum_j (1 + \Delta^2) \left(\frac{\partial^2 f_j}{\partial p_a \partial p_b} \right)^2. \quad \text{(A5)}$$

The covariance matrix, whose eigenvalues indicate the magnitude of the uncertainties, is given by

$$C_{ab}^{-1} = [\langle C_{ab} \rangle + \delta C_{ab}]^{-1} \approx \langle C_{ab} \rangle^{-1} [1 - \langle C_{ab} \rangle^{-1} \delta C_{ab}], \quad \text{(A6)}$$

where

$$\delta C_{ab} = \mu \sqrt{\sum_j (1 + \Delta^2) \left(\frac{\partial^2 f_j}{\partial p_a \partial p_b} \right)^2} \approx \mu \left(1 + \frac{\Delta^2}{2} \right) \sqrt{\sum_j \left(\frac{\partial^2 f_j}{\partial p_a \partial p_b} \right)^2} \quad \text{(A7)}$$

in which μ is a Gaussian random variable with unit variance. This error term is suppressed by approximately a factor of $N^{-1/2}$ relative to the mean variance, and thus in the limit of large N becomes insignificant.

The error term contains two elements, associated with the measurement and prediction errors, respectively. The ratio of the latter to the former is $\Delta^2/2$. When Δ^2 is small and N is large, the uncertainty on the parameter estimates then grows by a multiplicative factor of $\Delta^2/4$. This is unconditionally small when Δ is small.

In THEMIS, we typically set $\Delta = 0.25$, for which $\Delta^2/4 = 0.016$, which broadens the posterior parameter distributions by 1.6%. This does complicate the interpretation of fit quality for $N \gtrsim 2048$, however.

Appendix B
Error Distributions of Quantities Associated with Visibility Amplitudes

THEMIS has three data types associated with visibility amplitudes: the visibility amplitudes themselves, interferometric polarization fractions, and closure amplitudes. None of the underlying error distributions for any of these is Gaussian, and the latter two are poorly approximated by Gaussians. Here, we summarize what the relevant error distributions are and quantify how well they are approximated in THEMIS. In all cases, we will assume that the complex visibilities are well described by a Gaussian random variable with nonzero mean.

B.1. Visibility Amplitudes

The probability distribution of the magnitude of a complex Gaussian random variable, V, with mean V_0 and standard deviation σ, is given by the Rice distribution (see, e.g., Thompson et al. 2017):

$$p_r(|V|; |V_0|, \sigma) = \frac{|V|}{\sigma^2} e^{-(|V|^2 + |V_0|^2)/2\sigma^2} I_0 \left(\frac{|V||V_0|}{\sigma^2} \right). \quad \text{(B1)}$$

At high S/N (defined here by $|V_0|/\sigma$), a Gaussian with mean $\sqrt{|V_0|^2 + \sigma^2}$ and standard deviation σ becomes an increasingly good proxy for the Rice distribution, with the quality of this approximation increasing with S/N. When S/N $\geqslant 2$, the biased Gaussian is within 8% of the maximum probability of the Rice distribution for all values of $|V|$. These are compared in the left panel of Figure B1 for various choices of S/N. We provide a set of quantitative estimates of the accuracy of the Gaussian approximation in Table B1 for various S/Ns.

B.2. Visibility Amplitude Products

Before discussing the data quantities of interest, we begin by considering the distribution of the product of visibility amplitudes, i.e., $W = |V_A||V_B|$. We do this both to illustrate the procedure by which we construct exact probability distributions for combinations of products and quotients of visibility amplitudes and to show explicitly that these are typically well approximated by a single Rice distribution.

We begin by exploiting the nonnegative behavior of $|V|$ to define $v = \log(|V|)$. This simplifies the construction of the product by reducing it to a sum, i.e., in terms of v_A and v_B, $W = e^{v_A + v_B} \equiv e^w$. The probability distribution of v is given in terms of the Rice distribution by

$$q_r(v; |V_0|, \sigma) = e^v p_r(e^v; |V_0|, \sigma), \quad \text{(B2)}$$

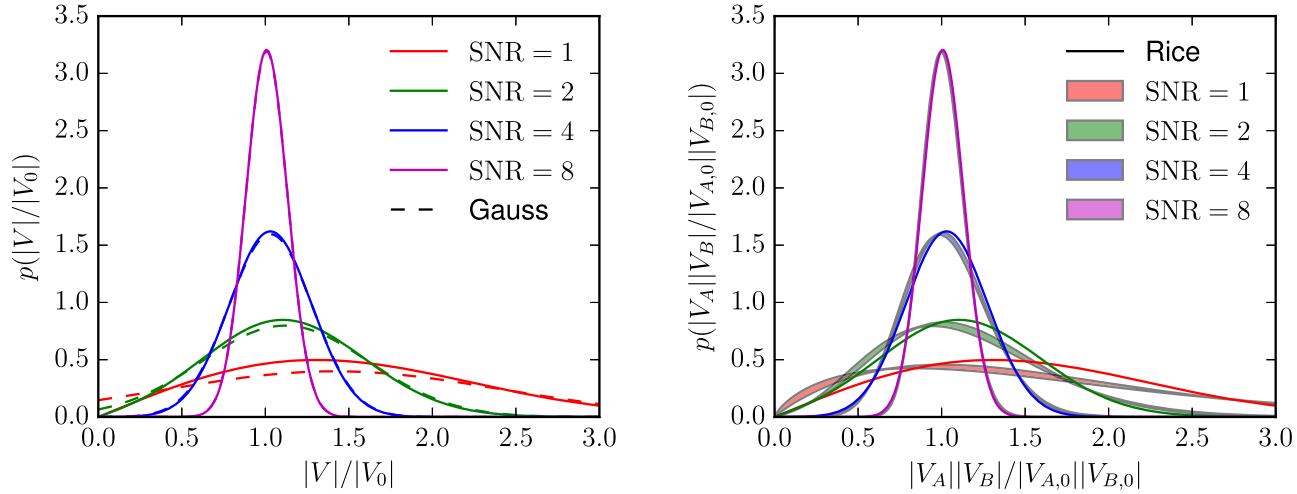

Figure B1. Left: comparison of the Rice and Gaussian distributions at various S/Ns. Right: distribution of products of visibility amplitudes and their comparison to a single Rice distribution. The S/N indicates $\sigma/W = \sqrt{\sigma_A^2/|V_{A,0}|^2 + \sigma_B^2/|V_{B,0}|^2}$, with the range for each S/N value corresponding to that obtained from different ways of apportioning the errors between $\sigma_A/|V_{A,0}|$ and $\sigma_B/|V_{B,0}|$. Summaries of the accuracy of the comparisons shown can be found in Tables B1 and B2.

Table B1
Accuracy of Visibility Amplitude Error Distribution Approximations

			S/N				
Approx.	Notes	Error	1	2	4	8	32
Gauss	…	$\delta_{\max}$ [a]	29%	8%	2%	0.4%	0.02%
…	…	δ_{mode} [b]	7%	1%	<0.1%	<0.1%	<0.1%
…	…	δ_w [c]	15%/17%	5%/6%	2%/2%	<0.5%/0.5%	<2%/<1%

Notes.
[a] Maximum absolute difference, measured relative to probability maximum.
[b] Fractional error in the location of the mode.
[c] Fractional error in the width of the region containing 68%/95% of the cumulative probability.

which we will call the logarithmic Rice distribution. Its characteristic function is

$$\phi_r(k;\,|V_0|,\,\sigma) = \int_{-\infty}^{\infty} e^{-ikv} q_r(v;\,|V_0|,\,\sigma). \tag{B3}$$

In practice, this may be computed efficiently via FFT. In terms of ϕ_r, the characteristic function of the probability distribution of w is, $\phi_p(k;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = \phi_r(k;\,|V_{A,0}|,\,\sigma_A)\,\phi_r(k;\,|V_{B,0}|,\,\sigma_B)$, and thus, the probability distribution of w is

$$q_p(w;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikw}\phi_r(k;\,|V_{A,0}|,\,\sigma_A)\,\phi_r(k;\,|V_{B,0}|,\,\sigma_B), \tag{B4}$$

which again may be computed efficiently via FFT. Finally, the desired probability distribution of W is then

$$p_p(W;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = W^{-1} q_p(\log W;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B). \tag{B5}$$

These are shown in Figure B1.

While formally the distribution of W is characterized by four parameters, in practice it is well approximated by a single Rice distribution with $W_0 = |V_{A,0}||V_{B,0}|$ and $\sigma = \sqrt{\sigma_A^2 + \sigma_B^2}$, differing by 17% of the maximum probability for S/N $\geqslant 2$. This comparison is also shown in the right panel of Figure B1, and estimates of the accuracy of the approximation are tabulated in Table B2 for various S/Ns. Note that this also implies that the product distribution is well fit by a Gaussian for sufficiently high S/N.

B.3. Polarization Fractions—Visibility Amplitude Quotients

We follow a similar procedure to that in the previous section to compute the distribution of the quotient of visibility amplitudes, i.e., $Q = |V_A|/|V_B|$. Unfortunately, we find that the remarkable simplicity of the distribution of visibility products does not extend to quotients. The characteristic function of the logarithmic quotient distribution is $\phi_q(k;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = \phi_r(k;\,|V_{A,0}|,\,\sigma_A)\,\phi_r{}^*(k;\,|V_{B,0}|,\,\sigma_B)$, where the $*$ denotes complex conjugation. The resulting logarithmic quotient probability distributions is

$$q_q(q;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikw}\phi_r(k;\,|V_{A,0}|,\,\sigma_A)\,\phi_r^*(k;\,|V_{B,0}|,\,\sigma_B), \tag{B6}$$

with the corresponding quotient probability distribution

$$p_q(Q;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B) = Q^{-1} q_q(\log Q;\,|V_{A,0}|,\,\sigma_A,\,|V_{B,0}|,\,\sigma_B). \tag{B7}$$

This is directly applicable to the polarization fraction, for which this distribution is shown in the left panel of Figure B2.

The polarization fraction distribution clearly deviates from the Gaussian and Rice distributions in two key respects. First, even at high S/N, the distributions are asymmetric, with the probability maximum lying below $Q_0 = |V_{A,0}|/|V_{B,0}|$. Second, there is a significant tail extending to high values of Q, containing sufficient weight to move the average Q above Q_0 for $S/N \geqslant 2$. More accurate are fitted (i.e., same mean and standard deviation) log-normal approximations, shown in the left panel of Figure B2, which recover the asymmetry, though still exhibit deviations for $S/N \leqslant 4$.

Combined with the accuracy of the Gaussian approximation to the Rice distribution, this motivates an exploration of better approximates to the quotient distribution of visibility amplitudes. For two Gaussian variables, with nonzero means, it is possible to analytically construct the quotient distribution analytically:

$$p_{\text{Gauss Q}}(r) = \int dxdy \frac{e^{-(x-\bar{x})/2\sigma_x^2}}{\sqrt{2\pi}\,\sigma_x} \frac{e^{-(y-\bar{y})^2/2\sigma_y^2}}{\sqrt{2\pi}\,\sigma_y} \delta \times \left(r - \frac{x}{y}\right) = \frac{e^{-(r-r_0)^2/2\Sigma^2}}{\sqrt{2\pi}\,\Sigma} \times \left[\Delta\,\mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\,\Omega}\right) - \frac{2\Omega}{\sqrt{2\pi}} e^{-\Delta^2/2\Omega^2}\right]$$

$$\text{where} \quad r_0 \equiv \frac{\bar{x}}{\bar{y}},\ \Sigma^2 \equiv r_0^2 \frac{\sigma_x^2}{\bar{x}^2} + r^2 \frac{\sigma_y^2}{\bar{y}^2},$$

$$\Delta \equiv \Sigma^{-2}\left(\frac{\sigma_x^2}{\bar{x}^2} + \frac{r}{r_0}\frac{\sigma_y^2}{\bar{y}^2}\right),\ \Omega \equiv \frac{\sigma_x \sigma_y}{\Sigma \bar{x} \bar{y}}, \tag{B8}$$

and $\mathrm{erf}(x)$ is the standard error function. This is also shown in the left panel of Figure B2. While deviations from p_q continue to exist, the Gaussian quotient (Gauss Q) distribution accurately reproduces the large high-Q tail.

As with $p_p(W)$, $p_q(Q)$ depends on the S/N of both the numerator and denominator independently. Unlike the product distribution, the quotient distribution is not symmetric in this dependency, with the properties of the denominator controlling the asymmetry and tail. Therefore, characterizing this distribution by a single pair of numbers—a central value and width—will result in a substantial uncertainty in the resulting quotient distribution. This is simplified for polarization fractions by the fact that the stations used to construct the visibility amplitudes in the numerator and denominator are the same, and thus both quantities have similar noise profiles in principle, i.e., $\sigma_A \approx \sigma_B$. This implies that

$$\frac{\sigma_A}{|V_{A,0}|} \approx \frac{1}{\sqrt{1+\breve{m}_0^2}} \frac{\sigma_{\breve{m}}}{\breve{m}_0} \quad \text{and} \quad \frac{\sigma_B}{|V_{B,0}|} \approx \frac{\breve{m}_0}{\sqrt{1+\breve{m}_0^2}} \frac{\sigma_{\breve{m}}}{\breve{m}_0}, \tag{B9}$$

where $\breve{m}_0 = |V_{A,0}|/|V_{B,0}|$ and $\sigma_{\breve{m}}$ is the uncertainty obtained by the standard error propagation formula. Note that these are the only two quantities required to fully specify the Gauss Q and quotient distributions. Combining this with Equation (B8), we obtain for the polarization fraction

$$p_{\breve{m}}(\breve{m};\, \breve{m}_0,\, \sigma_{\breve{m}}) = \frac{e^{-(\breve{m}-\breve{m}_0)^2/2\Sigma_{\breve{m}}^2}}{\sqrt{2\pi}\,\Sigma_{\breve{m}}} \times \left[\Delta_{\breve{m}}\,\mathrm{erf}\left(\frac{\Delta_{\breve{m}}}{\sqrt{2}\,\Omega_{\breve{m}}}\right) - \frac{2\Omega_{\breve{m}}}{\sqrt{2\pi}} e^{-\Delta_{\breve{m}}^2/2\Omega_{\breve{m}}^2}\right]$$

$$\text{where} \quad \Sigma_{\breve{m}}^2 = \sigma_{\breve{m}}^2 \frac{1+\breve{m}^2}{1+\breve{m}_0^2},\ \Delta_{\breve{m}} = \frac{1+\breve{m}\breve{m}_0}{1+\breve{m}^2},$$

$$\Omega_{\breve{m}}^2 = \frac{\sigma_{\breve{m}}^2}{(1+\breve{m}^2)(1+\breve{m}_0^2)}. \tag{B10}$$

For various $\breve{m}_0/\sigma_{\breve{m}}$, the right panel of Figure B2 shows comparisons of the exact quotient and Gauss Q distributions assuming the errors in Equation (B9) for $\breve{m}_0 = 0.5$ and 2.0. The Gauss Q distribution is within 11% and 13% of the maximum probability of the exact quotient distribution, respectively, at all values of $\breve{m}$ when $S/N \geqslant 2$. The accuracy of the Gauss Q approximation for the polarization fraction distribution is tabulated for different S/Ns and $\breve{m}$ in Table B3.

B.4. Closure Amplitudes

We now turn to the problem of constructing the error distribution for closure amplitudes generally. Again, the exact expression can be constructed using the characteristic functions of the logarithmic Rice distributions:

$$\phi_a(k;\, |V_A|,\, \sigma_A,\, |V_B|,\, \sigma_B,\, |V_C|,\, \sigma_C,\, |V_D|,\, \sigma_D) = \phi_r(k;\, |V_{A,0}|,\, \sigma_A)\phi_r(k;\, |V_{B,0}|,\, \sigma_B)\phi_r^* \times (k;\, |V_{C,0}|,\, \sigma_C)\phi_r^*(k;\, |V_{D,0}|,\, \sigma_D), \tag{B11}$$

from which we obtain

$$q_a(a;\, |V_{A,0}|,\, \sigma_A\, ,...) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ika}\phi_a(k;\, |V_{A,0}|,\, \sigma_A\, ,...), \tag{B12}$$

and

$$p_{\mathcal{V}}(\mathcal{V};\, |V_{A,0}|,\, \sigma_A\, ,...) = \mathcal{V}^{-1} q_a(\log \mathcal{V};\, |V_{A,0}|,\, \sigma_A\, ,...). \tag{B13}$$

These are shown in the left panel of Figure B3. Similar to the polarization fractions, they are clearly asymmetric and exhibit large tails to high values, typical of quotient distributions.

Formally, this requires knowledge of eight values to define. However, again, it is possible to accurately estimate $p_{\mathcal{V}}$ with only a handful of combinations of these values. Due to the similarity between the amplitude product distribution (Appendix B.2) and the Rice distribution, both the numerator and the denominator can be effectively described by only two parameters each. As a result, the closure amplitude distribution is similar to the amplitude quotient distribution described in Appendix B.3, shown by the dashed lines in the left panel of Figure B3.

If the S/Ns of the denominator and numerator are independently known, this is well approximated by the Gauss Q distribution in Equation (B8). These may be reconstructed with knowledge of the total S/N ($\mathcal{V}/\sigma_{\mathcal{V}}$), and the ratio of the thermal uncertainties in the numerator and denominator, i.e., $\sigma_n = \sqrt{\sigma_A^2 + \sigma_B^2}$ and $\sigma_d = \sqrt{\sigma_C^2 + \sigma_D^2}$, also shown in the left panel of Figure B3. The latter are not independent, related by

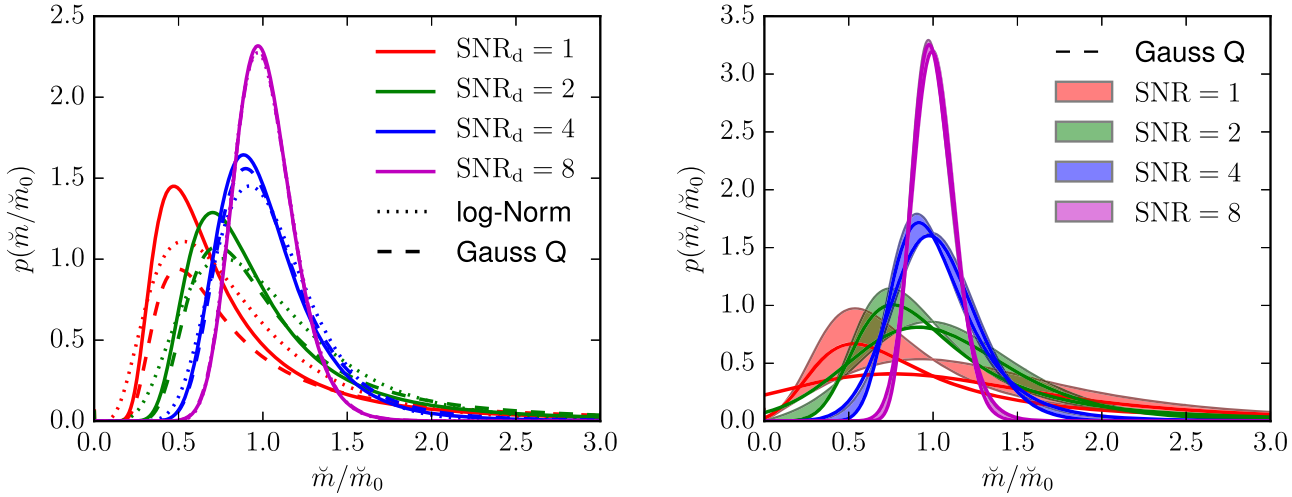

Figure B2. Left: comparison of the polarization fraction distribution (visibility amplitude quotient distribution) with the Gaussian quotient distribution approximate (Gauss Q) for various input values of the denominator S/N, $\mathrm{S/N}_d$. In all cases, the numerator S/N was set to 8. For comparison, a log-normal distribution is also shown. Right: comparison of the polarization fraction distribution (boundaries of the shaded region), the Gauss Q approximation in Equation (B10), and the log-normal model, for various total S/Ns (i.e., $\breve{m}/\sigma_{\breve{m}}$) at $\breve{m} = 0.5$ and 2.0. Quantitative estimates of the accuracy of the various approximations can be found in Table B3.

Table B2
Accuracy of Product Error Distribution Approximations

			S/N				
Approx.	Notes	Error	1	2	4	8	32
Rice	S/N ratio[a] = 1	$\delta_{\max}$[b]	33%	17%	9%	4%	1%
…	…	δ_{mode}[c]	36%	14%	4%	1%	<0.1%
…	…	δ_w[d]	16%/21%	7%/10%	2%/2%	0.6%/0.6%	<2%/<1%
Rice	S/N ratio = 4	$\delta_{\max}$	11%	5%	2%	1%	0.2%
…	…	δ_{mode}	13%	4%	1.0%	0.2%	<0.1%
…	…	δ_w	4%/9%	1%/3%	0.4%/0.7%	<0.5%/<0.2%	<2%/<1%

Notes.
[a] Ratio of $V_{A,0}/\sigma_A$ to $V_{B,0}/\sigma_B$ in the construction of the product distribution.
[b] Maximum absolute difference, measured relative to probability maximum.
[c] Fractional error in the location of the mode.
[d] Fractional error in the width of the region containing 68%/95% of the cumulative probability.

the repeated presence of each of the four stations required to produce a closure amplitude in the numerator and denominator. That is, identifying the baselines A, B, C, and D, with stations 1 and 2, 3 and 4, 1 and 4, and 2 and 3, respectively, assuming identical bandwidths and scan duration,

$$\rho^2 \equiv \frac{\sigma_n^2}{\sigma_d^2} = \frac{S_1S_2 + S_3S_4}{S_1S_4 + S_2S_3}, = \frac{1 + (S_3/S_1)(S_4/S_2)}{(S_4/S_2) + (S_3/S_1)}, \tag{B14}$$

where the S_j are station-specific system-equivalent flux densities (SEFDs). Despite the appearance of four SEFDs, this is a function of only two variables: the ratios S_3/S_1 and S_4/S_2. Where both of these ratios are of order unity, i.e., for a homogeneous array, $\rho \approx 1$. For highly heterogeneous arrays, in which more than one station is much more sensitive than the others, this can deviate from unity substantially, by an amount that depends on the second-lowest and second-highest SEFDs, $S_{\mathrm{2nd\,min}}$ and $S_{\mathrm{2nd\,max}}$, respectively,

$$\sqrt{\frac{S_{\mathrm{2nd\,min}}}{S_{\mathrm{2nd\,max}}}} \lesssim \rho \lesssim \sqrt{\frac{S_{\mathrm{2nd\,max}}}{S_{\mathrm{2nd\,min}}}}. \tag{B15}$$

For the 2017 EHT campaign, the station SEFDs ranged from 90 to 6000 Jy, with most being near 5000 Jy. The phased Atacama Large Millimeter/submillimeter Array (ALMA) is an extremely low-noise outlier at 90 Jy, followed by the Large Millimeter Telescope at 600 Jy (see the EHT imaging library, Chael et al. 2016, 2018b, and available at github.com/achael/eht-imaging). As a result, ρ ranges from 0.3 to 3.3, with most of the potential closure amplitude squares having ρ within 11% of unity. Therefore, even without prior knowledge about the visibilities that comprise the numerator and denominator of the closure amplitude, a similar procedure to that used for the closure amplitudes, where $\rho \approx 1$ is assumed, is well motivated.

Given a value of ρ, either from the station SEFDs or setting it to unity,

$$\frac{\sigma_n}{n} \approx \frac{\rho}{\sqrt{\rho^2 + \mathcal{V}_0^2}} \frac{\sigma_{\mathcal{V}}}{\mathcal{V}_0} \quad \text{and} \quad \frac{\sigma_d}{d} \approx \frac{\mathcal{V}_0}{\sqrt{\rho^2 + \mathcal{V}_0^2}} \frac{\sigma_{\mathcal{V}}}{\mathcal{V}_0}, \tag{B16}$$

where n and d are the mean values of $|V_A||V_B|$ and $|V_C||V_D|$, and $\mathcal{V}_0 = n/d$. As a result, an approximation of the closure amplitude distribution based on the Gauss Q distribution can

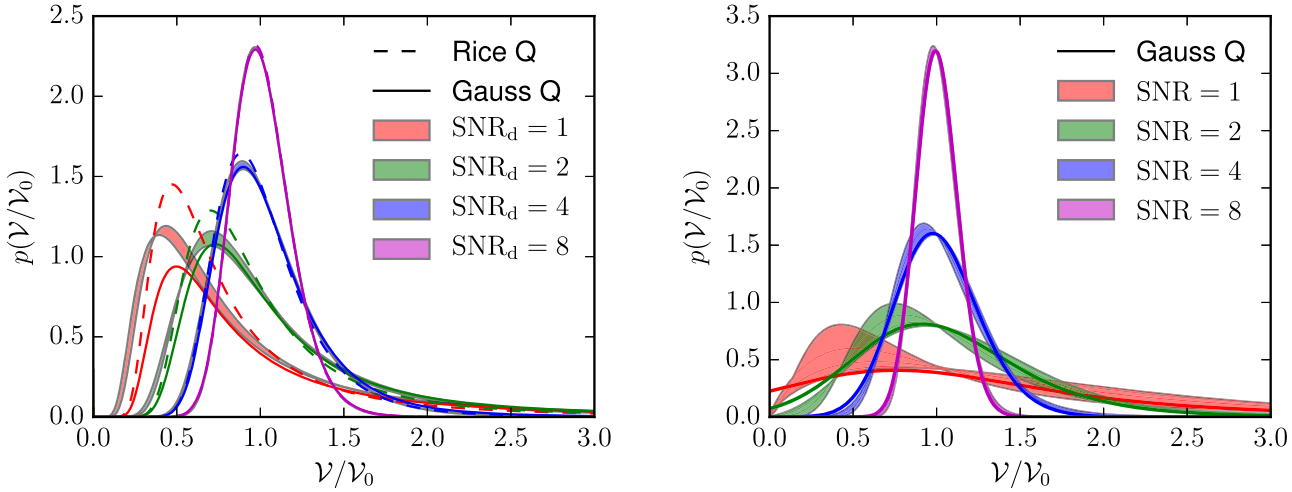

Figure B3. Left: comparison of the closure amplitude distribution with the Gauss Q approximation in Equation (B17) for various choices of the denominator S/N, $\mathrm{S/N}_d$. In all cases, the numerator S/N was set to 8, divided equally among the visibility amplitudes in the numerator. For comparison, the visibility amplitude quotient (Rice Q) distribution is also shown. The range of the filled bands indicates the uncertainty associated with various choices of how $\mathrm{S/N}_d$ is apportioned between the two visibility amplitudes in the denominator. In all cases, the value of ρ was set to the proper value in all models. Right: comparison of the closure amplitude distribution and the Gauss Q approximation with $\rho = 1$, for $\mathcal{V}_0 = 0.5$. The colored bands indicate the range of closure amplitude distributions when the true ρ is varied within the permissible range for the 2017 EHT campaign, [0.3,0.33]. The S/N within the numerator and denominator is distributed uniformly. Quantitative estimates of the accuracy of the various approximations for illustrative cases can be found in Table B4.

Table B3
Accuracy of Polarization Fraction Error Distribution Approximations

			S/N				
Approx.	Notes	Error	1	2	4	8	32
Gauss Q	$\breve{m}_0 = 0.5$	$\delta_{\max}$ [a]	41%	11%	5%	2%	0.6%
…	…	δ_{mode} [b]	22%	7%	2%	0.5%	<0.1%
…	…	δ_w [c]	3%/13%	5%/4%	2%/2%	<0.5%/0.4%	<2%/<1%
Gauss Q	$\breve{m}_0 = 2.0$	$\delta_{\max}$	31%	13%	6%	3%	0.6%
…	…	δ_{mode}	2%	2%	1%	0.4%	<0.1%
…	…	δ_w	17%/20%	13%/18%	5%/6%	1%/1%	<2%/<1%
log-norm	$\breve{m}_0 = 0.5$	$\delta_{\max}$	41%	17%	10%	5%	1%
…	…	δ_{mode}	26%	0.2%	0.4%	<0.1%	<0.1%
…	…	δ_w	46%/27%	21%/29%	5%/6%	1%/2%	<2%/<1%
log-norm	$\breve{m}_0 = 2.0$	$\delta_{\max}$	68%	48%	27%	13%	3%
…	…	δ_{mode}	27%	23%	9%	3%	0.2%
…	…	δ_w	>100%/63%	51%/23%	12%/3%	3%/0.6%	<2%/<1%

Notes.
[a] Maximum absolute difference, measured relative to probability maximum.
[b] Fractional error in the location of the mode.
[c] Fractional error in the width of the region containing 68%/95% of the cumulative probability.

be constructed via

$$p_{\mathcal{V}}(\mathcal{V};\mathcal{V}_0,\sigma_{\mathcal{V}}) = \frac{e^{-(\mathcal{V}-\mathcal{V}_0)^2/2\Sigma_{\mathcal{V}}^2}}{\sqrt{2\pi}\,\Sigma_{\mathcal{V}}} \times\left[\Delta_{\mathcal{V}}\,\mathrm{erf}\left(\frac{\Delta_{\mathcal{V}}}{\sqrt{2}\,\Omega_{\mathcal{V}}}\right) - \frac{2\Omega_{\mathcal{V}}}{\sqrt{2\pi}}e^{-\Delta_{\mathcal{V}}^2/2\Omega_{\mathcal{V}}^2}\right]$$

$$\text{where}\quad \Sigma_{\mathcal{V}}^2 = \sigma_{\mathcal{V}}^2\frac{\rho^2+\mathcal{V}^2}{\rho^2+\mathcal{V}_0^2},\quad \Delta_{\mathcal{V}} = \frac{\rho^2+\mathcal{V}\mathcal{V}_0}{\rho^2+\mathcal{V}^2},$$

$$\Omega_{\mathcal{V}}^2 = \frac{\rho^2\sigma_{\mathcal{V}}^2}{(\rho^2+\mathcal{V}^2)(\rho^2+\mathcal{V}_0^2)}. \tag{B17}$$

Quantitative assessments of the performance of this approximation for a number of variations in the distribution of S/Ns among the various components and for fixed and known ρ are listed in Table B4. In the Gauss Q cases with fixed ρ and for the log-normal distribution, we permit the true value of ρ to range from 0.3 to 3.3, reporting the maximum deviation for each measure independently.

Knowledge of ρ significantly improves the quality of the approximation, which for $\mathrm{S/N} \geqslant 2$ is accurate to 13% for $\mathcal{V} < 1$. The asymmetric impact of noise in the denominator and numerator of the closure amplitude is responsible for the worsening performance of the approximation when $\mathcal{V} > 1$; generally, closure amplitudes can be constructed so that $\mathcal{V} \lesssim 1$.

Table B4
Accuracy of Closure Amplitude Error Distribution Approximations

			S/N				
Approx.	Notes	Error	1	2	4	8	32
Gauss Q	$\mathcal{V}_0 = 0.5$,	$\delta_{\max}$[a]	54%	28%	13%	6%	2%
$\rho = 1$[b]	$S/N_{n,d}$ ratios[c] $= 0.5$,	$\delta_{\rm mode}$[d]	38%	17%	6%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$[e]	δ_w[f]	15%/14%	14%/11%	6%/8%	2%/4%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 0.5$,	$\delta_{\max}$	54%	28%	13%	6%	2%
$\rho = 1$	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	38%	17%	6%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$	δ_w	15%/14%	14%/11%	6%/8%	2%/4%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 2.0$,	$\delta_{\max}$	47%	25%	13%	6%	2%
$\rho = 1$	$S/N_{n,d}$ ratios $= 0.5$,	$\delta_{\rm mode}$	25%	17%	6%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$	δ_w	15%/5%	5%/26%	4%/15%	2%/5%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 2.0$,	$\delta_{\max}$	47%	25%	13%	6%	2%
$\rho = 1$	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	25%	17%	6%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$	δ_w	15%/5%	5%/26%	4%/15%	2%/5%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 0.5$,	$\delta_{\max}$	53%	13%	6%	2%	0.6%
ρ set[g]	$S/N_{n,d}$ ratios $= 0.5$,	$\delta_{\rm mode}$	5%	0.6%	<0.1%	<0.1%	<0.1%
⋯	$\rho = 3.3$	δ_w	9%/0.6%	0.4%/5%	0.5%/0.1%	<0.5%/<0.2%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 0.5$,	$\delta_{\max}$	53%	13%	6%	2%	0.6%
ρ set	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	5%	0.6%	<0.1%	<0.1%	<0.1%
⋯	$\rho = 3.3$	δ_w	9%/0.6%	0.4%/5%	0.5%/0.1%	<0.5%/<0.2%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 2.0$,	$\delta_{\max}$	46%	20%	8%	3%	0.6%
ρ set	$S/N_{n,d}$ ratios $= 0.5$,	$\delta_{\rm mode}$	15%	3%	0.4%	<0.1%	<0.1%
⋯	$\rho = 0.3$	δ_w	5%/6%	5%/16%	3%/10%	0.9%/2%	<2%/<1%
Gauss Q	$\mathcal{V}_0 = 2.0$,	$\delta_{\max}$	46%	20%	8%	3%	0.6%
ρ set	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	15%	3%	0.4%	<0.1%	<0.1%
⋯	$\rho = 0.3$	δ_w	5%/6%	5%/16%	3%/10%	0.9%/2%	<2%/<1%
log-norm	$\mathcal{V} = 0.5$,	$\delta_{\max}$	64%	43%	23%	11%	3%
⋯	$S/N_{n,d}$ ratios $= 0.5$,	$\delta_{\rm mode}$	36%	22%	8%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$	δ_w	62%/28%	32%/38%	10%/13%	3%/5%	<2%/<1%
log-norm	$\mathcal{V} = 0.5$,	$\delta_{\max}$	64%	43%	23%	11%	3%
⋯	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	36%	22%	8%	2%	<0.1%
⋯	$\rho \in [0.3, 3.3]$	δ_w	62%/28%	32%/38%	10%/13%	3%/5%	<2%/<1%
log-norm	$\mathcal{V} = 2.0$,	$\delta_{\max}$	76%	52%	29%	14%	3%
⋯	$S/N_{n,d}$ ratios $= 0.5$,	$\delta_{\rm mode}$	42%	27%	10%	3%	0.2%
⋯	$\rho \in [0.3, 3.3]$	δ_w	>100%/40%	41%/32%	12%/11%	4%/5%	<2%/<1%
log-norm	$\mathcal{V} = 2.0$,	$\delta_{\max}$	76%	52%	29%	14%	3%
⋯	$S/N_{n,d}$ ratios $= 2$,	$\delta_{\rm mode}$	42%	27%	10%	3%	0.2%
⋯	$\rho \in [0.3, 3.3]$	δ_w	>100%/40%	41%/32%	12%/11%	4%/5%	<2%/<1%

Notes.
[a] Maximum absolute difference, measured relative to probability maximum.
[b] Gauss Q model with ρ fixed at unity.
[c] Ratio of $V_{A,0}/\sigma_A$ to $V_{B,0}/\sigma_B$ and $V_{C,0}/\sigma_C$ to $V_{D,0}/\sigma_D$ in the construction of the closure amplitude distribution.
[d] Fractional error in the location of the mode.
[e] Range of input values of ρ explored in accuracy estimate.
[f] Fractional error in the width of the region containing 68%/95% of the cumulative probability.
[g] Gauss Q model with ρ set to input value.

The performance of the approximation in reconstructing the mode and width of the distribution is very good in this limit, better than 1%. When ρ is not known a priori, setting it to unity introduces an additional error in the approximation of the closure amplitude distribution. Nevertheless, even with excursions of a factor of 3, by $S/N \geqslant 4$, the Gauss Q approximation is accurate to 13% at all $\mathcal{V}$.

In practice, the primary difficulty with applying Equation (B17) is the accuracy with which σ_n/n and σ_d/d can be reconstructed, which depends how close $\mathcal{V}$ is to $\mathcal{V}_0$. At low S/N, this can lead to a significant error in the estimation of the likelihood. Where the estimate of σ_n/n or σ_d/d is higher than their true values, this makes little difference. However, where they are much lower than their true values, this can result

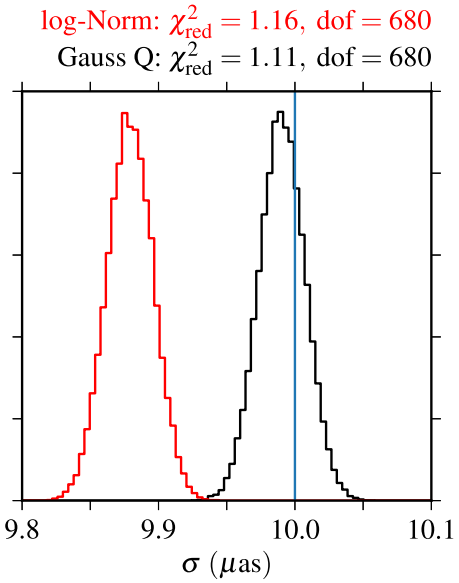

Figure B4. Comparison of the posterior distribution of the reconstructed size of a Gaussian from simulated closure amplitude data assuming the Gauss $\mathcal{Q}$ (black) and log-normal approximations (red). The true size is 10 μas, indicated by the blue line. An S/N cut of S/N > 4 was applied to the data to exclude marginal detections and improve the fidelity of the assumed approximate error distributions.

in a distribution that is considerably more narrowly concentrated about $\mathcal{V}_0$ than the true distribution, biasing any resulting parameter estimates. This ceases to be a significant bias for S/N > 4 for the 2017 campaign.

Finally, we remark on comparisons to a frequent alternative approximation for the closure amplitude error distribution, the log-normal distribution. This has a number of desirable features: it is simple to define and rapid to compute, it eliminates the conceptual distribution between the numerator and denominator in the definition of the closure amplitudes, and it naturally produces extended tails toward large values. A quantitative comparison to the Gauss $\mathcal{Q}$ approximation is presented in Table B4. To do so, we have constructed a biased log-normal distribution, i.e., with mean $\log(\mathcal{V}_0 + \sigma_{\mathcal{V}}^2/\mathcal{V}_0)$ and standard deviation $\sigma_{\mathcal{V}}/\mathcal{V}_0$. These perform more poorly than the Gauss $\mathcal{Q}$ approximation at all values of S/N, exhibiting significant biases in the maximum and median of the error distributions.

At high S/Ns, this distinction makes little difference; for modest S/Ns, this results in significant systematic biases in reconstructed structural parameters. This is clear in Figure B4, which presents the posterior distributions of the size of a Gaussian feature from simulated 2017 campaign data. The underlying image and thermal noise were chosen such that the S/Ns of the closure amplitudes are moderate, i.e., they range from below our cutoff of 4–18. The two posteriors shown are for two likelihoods that differ in the assumed closure amplitude error distribution. For both, excellent fits are obtained. However, in the latter case, when a log-normal distribution is assumed, a notable and systematic bias toward more compact structures is present. For this reason, we adopt the modestly more complicated Gauss $\mathcal{Q}$ distribution.

Appendix C
Gaussian-prior-modified Levenberg–Marquardt Algorithm

To numerically maximize the likelihood during the reconstruction and marginalization over station gain reconstructions (Section 6.8), we employ a modified Levenberg–Marquardt algorithm that includes Gaussian priors on the gains. We follow Section 15 of Press et al. (1992), in which algorithms to minimize a χ^2 are presented. The appropriately modified expression for the χ^2, accounting for a Gaussian prior on the parameters, is

$$\tilde{\chi}^2(\boldsymbol{g}) = \chi^2(\boldsymbol{g}) + \sum_A \frac{g_A^2}{\Sigma_A^2}, \tag{C1}$$

where $\boldsymbol{g}$, g_A, and Σ_A are all defined in Section 6.8, and the additional term is the Gaussian prior directly. The associated modified definitions of β_A and α_{AB} (defined in 15.5.8 of Press et al. 1992) are

$$\tilde{\beta}_A = -\frac{1}{2}\frac{\partial\tilde{\chi}^2}{\partial g_A} = \beta_A - \frac{g_A}{\Sigma_A^2} \tag{C2}$$

and

$$\tilde{\alpha}_{AB} = \frac{1}{2}\frac{\partial^2\tilde{\chi}^2}{\partial g_A \partial g_B} = \alpha_{AB} + \frac{\delta_{AB}}{\Sigma_A^2}. \tag{C3}$$

The expression for α_{AB} is simplified further in 15.5.11 of Press et al. (1992), though we continue to adopt the above modification. All other elements of the Levenberg–Marquardt algorithm are unchanged.

ORCID iDs

Avery E. Broderick https://orcid.org/0000-0002-3351-760X
Roman Gold https://orcid.org/0000-0003-2492-1966
Mansour Karami https://orcid.org/0000-0001-7387-9333
Jorge A. Preciado-López https://orcid.org/0000-0002-4146-0113
Paul Tiede https://orcid.org/0000-0003-3826-5648
Hung-Yi Pu https://orcid.org/0000-0001-9270-8812
Kazunori Akiyama https://orcid.org/0000-0002-9475-4254
Antxon Alberdi https://orcid.org/0000-0002-9371-1033
Rebecca Azulay https://orcid.org/0000-0002-2200-5393
Anne-Kathrin Baczko https://orcid.org/0000-0003-3090-3975
Mislav Baloković https://orcid.org/0000-0003-0476-6647
John Barrett https://orcid.org/0000-0002-9290-0764
Lindy Blackburn https://orcid.org/0000-0002-9030-642X
Katherine L. Bouman https://orcid.org/0000-0003-0077-4367
Geoffrey C. Bower https://orcid.org/0000-0003-4056-9982
Christiaan D. Brinkerink https://orcid.org/0000-0002-2322-0749
Roger Brissenden https://orcid.org/0000-0002-2556-0894
Silke Britzen https://orcid.org/0000-0001-9240-6734
Thomas Bronzwaer https://orcid.org/0000-0003-1151-3971
Do-Young Byun https://orcid.org/0000-0003-1157-4109
John E. Carlstrom https://orcid.org/0000-0002-2044-7665
Andrew Chael https://orcid.org/0000-0003-2966-6220
Shami Chatterjee https://orcid.org/0000-0002-2878-1502
Koushik Chatterjee https://orcid.org/0000-0002-2825-3590
Ming-Tang Chen https://orcid.org/0000-0001-6573-3318
Yongjun Chen (陈永军) https://orcid.org/0000-0001-5650-6770
Ilje Cho https://orcid.org/0000-0001-6083-7521
John E. Conway https://orcid.org/0000-0003-2448-9181

James M. Cordes https://orcid.org/0000-0002-4049-1882
Geoffrey B. Crew https://orcid.org/0000-0002-2079-3189
Yuzhu Cui https://orcid.org/0000-0001-6311-4345
Jordy Davelaar https://orcid.org/0000-0002-2685-2434
Mariafelicia De Laurentis https://orcid.org/0000-0002-9945-682X
Roger Deane https://orcid.org/0000-0003-1027-5043
Jessica Dempsey https://orcid.org/0000-0003-1269-9667
Gregory Desvignes https://orcid.org/0000-0003-3922-4055
Sheperd S. Doeleman https://orcid.org/0000-0002-9031-0904
Ralph P. Eatough https://orcid.org/0000-0001-6196-4135
Heino Falcke https://orcid.org/0000-0002-2526-6724
Vincent L. Fish https://orcid.org/0000-0002-7128-9345
Ed Fomalont https://orcid.org/0000-0002-9036-2747
Raquel Fraga-Encinas https://orcid.org/0000-0002-5222-1361
Per Friberg https://orcid.org/0000-0002-8010-8454
Christian M. Fromm https://orcid.org/0000-0002-1827-1656
Peter Galison https://orcid.org/0000-0002-6429-3872
Charles F. Gammie https://orcid.org/0000-0001-7451-8935
Roberto García https://orcid.org/0000-0002-6584-7443
Boris Georgiev https://orcid.org/0000-0002-3586-6424
Ciriaco Goddi https://orcid.org/0000-0002-2542-7743
José L. Gómez https://orcid.org/0000-0003-4190-7613
Minfeng Gu (顾敏峰) https://orcid.org/0000-0002-4455-6946
Mark Gurwell https://orcid.org/0000-0003-0685-3621
Kazuhiro Hada https://orcid.org/0000-0001-6906-772X
Ronald Hesper https://orcid.org/0000-0003-1918-6098
Luis C. Ho (何子山) https://orcid.org/0000-0001-6947-5846
Paul Ho https://orcid.org/0000-0002-3412-4306
Mareki Honma https://orcid.org/0000-0003-4058-9000
Lei Huang (黄磊) https://orcid.org/0000-0002-1923-227X
Makoto Inoue https://orcid.org/0000-0001-5037-3989
Sara Issaoun https://orcid.org/0000-0002-5297-921X
David J. James https://orcid.org/0000-0001-5160-4486
Michael Janssen https://orcid.org/0000-0001-8685-6544
Britton Jeter https://orcid.org/0000-0003-2847-1712
Wu Jiang (江悟) https://orcid.org/0000-0001-7369-3539
Alejandra Jiménez-Rosales https://orcid.org/0000-0002-2662-3754
Michael D. Johnson https://orcid.org/0000-0002-4120-3029
Svetlana Jorstad https://orcid.org/0000-0001-6158-1708
Taehyun Jung https://orcid.org/0000-0001-7003-8643
Ramesh Karuppusamy https://orcid.org/0000-0002-5307-2919
Tomohisa Kawashima https://orcid.org/0000-0001-8527-0496
Garrett K. Keating https://orcid.org/0000-0002-3490-146X
Mark Kettenis https://orcid.org/0000-0002-6156-5617
Jae-Young Kim https://orcid.org/0000-0001-8229-7183
Jongsoo Kim https://orcid.org/0000-0002-1229-0426
Motoki Kino https://orcid.org/0000-0002-2709-7338
Jun Yi Koay https://orcid.org/0000-0002-7029-6658
Patrick M. Koch https://orcid.org/0000-0003-2777-5861
Shoko Koyama https://orcid.org/0000-0002-3723-3372
Michael Kramer https://orcid.org/0000-0002-4175-2271
Carsten Kramer https://orcid.org/0000-0002-4908-4925
Thomas P. Krichbaum https://orcid.org/0000-0002-4892-9586
Cheng-Yu Kuo https://orcid.org/0000-0001-6211-5581
Sang-Sung Lee https://orcid.org/0000-0002-6269-594X
Yan-Rong Li (李彦荣) https://orcid.org/0000-0001-5841-9179
Zhiyuan Li (李志远) https://orcid.org/0000-0003-0355-6437
Michael Lindqvist https://orcid.org/0000-0002-3669-0715
Rocco Lico https://orcid.org/0000-0001-7361-2460
Kuo Liu https://orcid.org/0000-0002-2953-7376
Elisabetta Liuzzo https://orcid.org/0000-0003-0995-5201
Wen-Ping Lo https://orcid.org/0000-0003-1869-2503
Andrei P. Lobanov https://orcid.org/0000-0003-1622-1484
Laurent Loinard https://orcid.org/0000-0002-5635-3345
Colin Lonsdale https://orcid.org/0000-0003-4062-4654
Ru-Sen Lu (路如森) https://orcid.org/0000-0002-7692-7967
Nicholas R. MacDonald https://orcid.org/0000-0002-6684-8691
Jirong Mao (毛基荣) https://orcid.org/0000-0002-7077-7195
Alan P. Marscher https://orcid.org/0000-0001-7396-3332
Iván Martí-Vidal https://orcid.org/0000-0003-3708-9611
Lynn D. Matthews https://orcid.org/0000-0002-3728-8082
Karl M. Menten https://orcid.org/0000-0001-6459-0669
Yosuke Mizuno https://orcid.org/0000-0002-8131-6730
Izumi Mizuno https://orcid.org/0000-0002-7210-6264
James M. Moran https://orcid.org/0000-0002-3882-4414
Kotaro Moriyama https://orcid.org/0000-0003-1364-3761
Monika Moscibrodzka https://orcid.org/0000-0002-4661-6332
Cornelia Müller https://orcid.org/0000-0002-2739-2994
Hiroshi Nagai https://orcid.org/0000-0003-0292-3645
Neil M. Nagar https://orcid.org/0000-0001-6920-662X
Masanori Nakamura https://orcid.org/0000-0001-6081-2420
Ramesh Narayan https://orcid.org/0000-0002-1919-2730
Gopal Narayanan https://orcid.org/0000-0002-4723-6569
Iniyan Natarajan https://orcid.org/0000-0001-8242-4373
Roberto Neri https://orcid.org/0000-0002-7176-4046
Chunchong Ni https://orcid.org/0000-0003-1361-5699
Hiroki Okino https://orcid.org/0000-0003-3779-2016
Héctor Olivares https://orcid.org/0000-0001-6833-7580
Gisela N. Ortiz-León https://orcid.org/0000-0002-2863-676X
Tomoaki Oyama https://orcid.org/0000-0003-4046-2923
Daniel C. M. Palumbo https://orcid.org/0000-0002-7179-3816
Jongho Park https://orcid.org/0000-0001-6558-9053
Ue-Li Pen https://orcid.org/0000-0003-2155-9578
Dominic W. Pesce https://orcid.org/0000-0002-5278-9221
Richard Plambeck https://orcid.org/0000-0001-6765-9609
Oliver Porth https://orcid.org/0000-0002-4584-2557
Ben Prather https://orcid.org/0000-0002-0393-7734
Venkatessh Ramakrishnan https://orcid.org/0000-0002-9248-086X
Ramprasad Rao https://orcid.org/0000-0002-1407-7944
Mark G. Rawlings https://orcid.org/0000-0002-6529-202X
Alexander W. Raymond https://orcid.org/0000-0002-5779-4767

Luciano Rezzolla https://orcid.org/0000-0002-1330-7103
Bart Ripperda https://orcid.org/0000-0002-7301-3908
Freek Roelofs https://orcid.org/0000-0001-5461-3687
Alan Rogers https://orcid.org/0000-0003-1941-7458
Eduardo Ros https://orcid.org/0000-0001-9503-4892
Mel Rose https://orcid.org/0000-0002-2016-8746
Helge Rottmann https://orcid.org/0000-0003-1799-8228
Chet Ruszczyk https://orcid.org/0000-0001-7278-9707
Benjamin R. Ryan https://orcid.org/0000-0001-8939-4461
Kazi L. J. Rygl https://orcid.org/0000-0003-4146-9043
Salvador Sánchez https://orcid.org/0000-0002-8042-5951
David Sánchez-Arguelles https://orcid.org/0000-0002-7344-9920
Mahito Sasada https://orcid.org/0000-0001-5946-9960
Tuomas Savolainen https://orcid.org/0000-0001-6214-1085
Karl-Friedrich Schuster https://orcid.org/0000-0003-2890-9454
Lijing Shao https://orcid.org/0000-0002-1334-8853
Zhiqiang Shen (沈志强) https://orcid.org/0000-0003-3540-8746
Des Small https://orcid.org/0000-0003-3723-5404
Bong Won Sohn https://orcid.org/0000-0002-4148-8378
Jason SooHoo https://orcid.org/0000-0003-1938-0720
Fumie Tazaki https://orcid.org/0000-0003-0236-0600
Remo P. J. Tilanus https://orcid.org/0000-0002-6514-553X
Michael Titus https://orcid.org/0000-0002-3423-4505
Kenji Toma https://orcid.org/0000-0002-7114-6010
Pablo Torne https://orcid.org/0000-0001-8700-6058
Efthalia Traianou https://orcid.org/0000-0002-1209-6500
Sascha Trippe https://orcid.org/0000-0003-0465-1559
Ilse van Bemmel https://orcid.org/0000-0001-5473-2950
Huib Jan van Langevelde https://orcid.org/0000-0002-0230-5946
Daniel R. van Rossum https://orcid.org/0000-0001-7772-6131
Jan Wagner https://orcid.org/0000-0003-1105-6109
John Wardle https://orcid.org/0000-0002-8960-2942
Jonathan Weintroub https://orcid.org/0000-0002-4603-5204
Norbert Wex https://orcid.org/0000-0003-4058-2837
Robert Wharton https://orcid.org/0000-0002-7416-5209
Maciek Wielgus https://orcid.org/0000-0002-8635-4242
George N. Wong https://orcid.org/0000-0001-6952-2147
Qingwen Wu (吴庆文) https://orcid.org/0000-0003-4773-4987
Doosoo Yoon https://orcid.org/0000-0001-8694-8166
André Young https://orcid.org/0000-0003-0000-2682
Ken Young https://orcid.org/0000-0002-3666-4920
Ziri Younsi https://orcid.org/0000-0001-9283-1191
Feng Yuan (袁峰) https://orcid.org/0000-0003-3564-6437
Ye-Fei Yuan (袁业飞) https://orcid.org/0000-0002-7330-4756
J. Anton Zensus https://orcid.org/0000-0001-7470-3321
Guangyao Zhao https://orcid.org/0000-0002-4417-1659
Shan-Shan Zhao https://orcid.org/0000-0002-9774-3606
Ziyan Zhu https://orcid.org/0000-0003-4463-5828

References

Akiyama, K., Lu, R.-S., Fish, V. L., et al. 2015, ApJ, 807, 150
Bambi, C., & Freese, K. 2009, PhRvD, 79, 043002
Benkevitch, L., Akiyama, K., Lu, R., Doeleman, S., & Fish, V. 2016, arXiv:1609.00055
Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, ApJ, 894, 31
Bower, G. C., Broderick, A., Dexter, J., et al. 2018, ApJ, 868, 101
Bower, G. C., Goss, W. M., Falcke, H., Backer, D. C., & Lithwick, Y. 2006, ApJL, 648, L127
Braak, C. J. F. T. 2006, Stat. Comput., 16, 239
Broderick, A., & Blandford, R. 2003, MNRAS, 342, 1280
Broderick, A., & Blandford, R. 2004, MNRAS, 349, 994
Broderick, A. E., Fish, V. L., Doeleman, S. S., & Loeb, A. 2011, ApJ, 735, 110
Broderick, A. E., Fish, V. L., Johnson, M. D., et al. 2016, ApJ, 820, 137
Broderick, A. E., Johannsen, T., Loeb, A., & Psaltis, D. 2014, ApJ, 784, 7
Broderick, A. E., & Loeb, A. 2005, MNRAS, 363, 353
Broderick, A. E., & Loeb, A. 2006a, MNRAS, 367, 905
Broderick, A. E., & Loeb, A. 2006b, ApJL, 636, L109
Broderick, A. E., & Loeb, A. 2009a, ApJ, 697, 1164
Broderick, A. E., & Loeb, A. 2009b, ApJL, 703, L104
Broderick, A. E., & Tchekhovskoy, A. 2015, ApJ, 809, 97
Bromley, B. C., Melia, F., & Liu, S. 2001, ApJL, 555, L83
Bronzwaer, T., Davelaar, J., Younsi, Z., et al. 2018, A&A, 613, A2
Chael, A., Rowan, M., Narayan, R., Johnson, M., & Sironi, L. 2018a, MNRAS, 478, 5209
Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018b, ApJ, 857, 23
Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, ApJ, 829, 11
Chan, C.-K., Liu, S., Fryer, C. L., et al. 2009, ApJ, 701, 521
Chan, C.-k., Medeiros, L., Ozel, F., & Psaltis, D. 2018, ApJ, 867, 59
Chan, C.-K., Psaltis, D., Özel, F., Narayan, R., & Sądowski, A. 2015, ApJ, 799, 1
Chandra, M., Gammie, C. F., Foucart, F., & Quataert, E. 2015, ApJ, 810, 162
Chatterjee, K., Liska, M., Tchekhovskoy, A., & Markoff, S. B. 2019, MNRAS, 490, 2200
Cunningham, C. T. 1975, ApJ, 202, 788
Davelaar, J., Moscibrodzka, M., Bronzwaer, T., & Falcke, H. 2018, A&A, 612, A34
de Gasperin, F., Orrú, E., Murgia, M., et al. 2012, A&A, 547, A56
Dexter, J. 2016, MNRAS, 462, 115
Dexter, J., & Agol, E. 2009, ApJ, 696, 1616
Dexter, J., Agol, E., & Fragile, P. C. 2009, ApJL, 703, L142
Dexter, J., Agol, E., Fragile, P. C., & McKinney, J. C. 2010, ApJ, 717, 1092
Dexter, J., & Fragile, P. C. 2013, MNRAS, 432, 2252
Dexter, J., McKinney, J. C., & Agol, E. 2012, MNRAS, 421, 1517
Doeleman, S. 2010, in 10th European VLBI Network Symp. and EVN Users Meeting: VLBI and the New Generation of Radio Arrays (Trieste: SISSA), 53
Doeleman, S., Agol, E., Backer, D., et al. 2009, Astro2010: The Astronomy and Astrophysics Decadal Survey, Science White Paper No. 68
Doeleman, S. S., Fish, V. L., Schenck, D. E., et al. 2012, Sci, 338, 355
Doeleman, S. S., Weintroub, J., Rogers, A. E. E., et al. 2008, Natur, 455, 78
Dolence, J. C., Gammie, C. F., Mościbrodzka, M., & Leung, P. K. 2009, ApJS, 184, 387
Dolence, J. C., Gammie, C. F., Shiokawa, H., & Noble, S. C. 2012, ApJL, 746, L10
Eckart, A., Baganoff, F. K., Schödel, R., et al. 2006, A&A, 450, 535
Eckart, A., Schödel, R., García-Marín, M., et al. 2008, A&A, 492, 337
Event Horizon Telescope Collaboration et al. 2019a, ApJL, 875, L1
Event Horizon Telescope Collaboration et al. 2019b, ApJL, 875, L2
Event Horizon Telescope Collaboration et al. 2019c, ApJL, 875, L3
Event Horizon Telescope Collaboration et al. 2019d, ApJL, 875, L4
Event Horizon Telescope Collaboration et al. 2019e, ApJL, 875, L5
Event Horizon Telescope Collaboration et al. 2019f, ApJL, 875, L6
Fabian, A. C. 2012, ARA&A, 50, 455
Falcke, H. 2017, JPhCS, 942, 012001
Falcke, H., Goss, W. M., Matsuo, H., et al. 1998, ApJ, 499, 731
Falcke, H., & Markoff, S. 2000, A&A, 362, 113
Fish, V. L., Doeleman, S. S., Beaudoin, C., et al. 2011, ApJL, 727, L36
Fish, V. L., Doeleman, S. S., Broderick, A. E., Loeb, A., & Rogers, A. E. E. 2009, ApJ, 706, 1353
Fish, V. L., Johnson, M. D., Doeleman, S. S., et al. 2016, ApJ, 820, 90
Fraga-Encinas, R., Mościbrodzka, M., Brinkerink, C., & Falcke, H. 2016, A&A, 588, A57
Frigo, M., & Johnson, S. G. 2005, Proc. IEEE, 93, 216
Fuerst, S. V., & Wu, K. 2004, A&A, 424, 733
Gammie, C. F., & Leung, P. K. 2012, ApJ, 752, 123
Gammie, C. F., McKinney, J. C., & Tóth, G. 2003, ApJ, 589, 444
Gold, R., McKinney, J. C., Johnson, M. D., & Doeleman, S. S. 2017, ApJ, 837, 180
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Heckman, T. M., & Best, P. N. 2014, ARA&A, 52, 589
Herrnstein, J. R., Greenhill, L. J., Moran, J. M., et al. 1998, ApJL, 497, L69
Hilbert, D. 1917, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen—Mathematisch-physikalische Klasse (Berlin: Weidmannsche Buchhandlung), 53
Hirotani, K., & Pu, H.-Y. 2016, ApJ, 818, 50

Huang, L., & Shcherbakov, R. V. 2011, MNRAS, 416, 2574
Huang, L., Takahashi, R., & Shen, Z.-Q. 2009, ApJ, 706, 960
Huijser, D., Goodman, J., & Brewer, B. J. 2015, arXiv:1509.02230
Issaoun, S., Johnson, M. D., Blackburn, L., et al. 2019, ApJ, 871, 30
Jennison, R. C. 1958, MNRAS, 118, 276
Jeter, B., Broderick, A. E., & Gold, R. 2020, MNRAS, 493, 5606
Jiménez-Rosales, A., & Dexter, J. 2018, MNRAS, 478, 1875
Johannsen, T. 2013, PhRvD, 87, 124017
Johannsen, T., & Psaltis, D. 2010, ApJ, 718, 446
Johannsen, T., Wang, C., Broderick, A. E., et al. 2016, PhRvL, 117, 091101
Johnson, M. D., Fish, V. L., Doeleman, S. S., et al. 2014, ApJ, 794, 150
Johnson, M. D., Fish, V. L., Doeleman, S. S., et al. 2015, Sci, 350, 1242
Johnson, M. D., & Gwinn, C. R. 2015, ApJ, 805, 180
Johnson, M. D., & Narayan, R. 2016, ApJ, 826, 170
Johnson, M. D., Narayan, R., Psaltis, D., et al. 2018, ApJ, 865, 104
Kamruddin, A. B., & Dexter, J. 2013, MNRAS, 434, 765
Kim, J., Marrone, D. P., Chan, C.-K., et al. 2016, ApJ, 832, 156
Knuth, K. H., Habeck, M., Malakar, N. K., Mubeen, A. M., & Placek, B. 2015, DSP, 47, 50
Lartillot, N., & Philippe, H. 2006, Syst. Biol., 55, 195
Levinson, A., & Rieger, F. 2011, ApJ, 730, 123
Liska, M., Hesp, C., Tchekhovskoy, A., et al. 2018, MNRAS, 474, L81
Lu, R.-S., Krichbaum, T. P., Roy, A. L., et al. 2018, ApJ, 859, 60
Lu, R.-S., Roelofs, F., Fish, V. L., et al. 2016, ApJ, 817, 173
Luminet, J.-P. 1979, A&A, 75, 228
Mao, S. A., Dexter, J., & Quataert, E. 2017, MNRAS, 466, 4307
Marrone, D. P. 2006, PhD thesis, Harvard Univ.
McKinney, J. C., Tchekhovskoy, A., & Blandford, R. D. 2012, MNRAS, 423, 3083
McKinney, J. C., Tchekhovskoy, A., Sadowski, A., & Narayan, R. 2014, MNRAS, 441, 3177
Medeiros, L., Chan, C.-k., Özel, F., et al. 2017, ApJ, 844, 35
Medeiros, L., Chan, C.-k., Özel, F., et al. 2018, ApJ, 856, 163
Middelberg, E., Roy, A. L., Walker, R. C., & Falcke, H. 2005, A&A, 433, 897
Mizuno, Y., Younsi, Z., Fromm, C. M., et al. 2018, NatAs, 2, 585
Mościbrodzka, M., Dexter, J., Davelaar, J., & Falcke, H. 2017, MNRAS, 468, 2214
Mościbrodzka, M., Falcke, H., & Shiokawa, H. 2016, A&A, 586, A38
Mościbrodzka, M., Falcke, H., Shiokawa, H., & Gammie, C. F. 2014, A&A, 570, A7
Mościbrodzka, M., & Gammie, C. F. 2018, MNRAS, 475, 43
Mościbrodzka, M., Gammie, C. F., Dolence, J. C., Shiokawa, H., & Leung, P. K. 2009, ApJ, 706, 497
Muñoz, D. J., Marrone, D. P., Moran, J. M., & Rao, R. 2012, ApJ, 745, 115
Narayan, R., Mahadevan, R., Grindlay, J. E., Popham, R. G., & Gammie, C. 1998, ApJ, 492, 554
Natarajan, I., Paragi, Z., Zwart, J., et al. 2017, MNRAS, 464, 4306
Nelson, B., Ford, E. B., & Payne, M. J. 2014, ApJS, 210, 11
Noble, S. C., Gammie, C. F., McKinney, J. C., & del Zanna, L. 2006, ApJ, 641, 626
Olivares, H., Porth, O., & Mizuno, Y. 2018, arXiv:1802.00860
Özel, F., Psaltis, D., & Narayan, R. 2000, ApJ, 541, 234
Porth, O., Olivares, H., Mizuno, Y., et al. 2017, ComAC, 4, 1
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C. The Art of Scientific Computing (Cambridge: Cambridge Univ. Press)
Psaltis, D., Johnson, M., Narayan, R., et al. 2018, arXiv:1805.01242
Psaltis, D., Wex, N., & Kramer, M. 2016, ApJ, 818, 121
Pu, H.-Y., Akiyama, K., & Asada, K. 2016a, ApJ, 831, 4
Pu, H.-Y., & Broderick, A. E. 2018, ApJ, 863, 148
Pu, H.-Y., Yun, K., Younsi, Z., & Yoon, S.-J. 2016b, ApJ, 820, 105
Reid, M. J. 1999, in ASP Conf. Ser. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley (San Francisco, CA: ASP), 481
Ressler, S. M., Tchekhovskoy, A., Quataert, E., Chandra, M., & Gammie, C. F. 2015, MNRAS, 454, 1848
Ressler, S. M., Tchekhovskoy, A., Quataert, E., & Gammie, C. F. 2017, MNRAS, 467, 3604
Ricarte, A., & Dexter, J. 2015, MNRAS, 446, 1973
Roberts, D. H., Wardle, J. F. C., & Brown, L. F. 1994, ApJ, 427, 718
Roelofs, F., Johnson, M. D., Shiokawa, H., Doeleman, S. S., & Falcke, H. 2017, ApJ, 847, 55
Rogers, A. E. E., Moffet, A. T., Backer, D. C., & Moran, J. M. 1984, RaSc, 19, 1552
Shcherbakov, R. V. 2014, ASTRORAY: General Relativistic Polarized Radiative Transfer Code, Version 1.0, Astrophysics Source Code Library, ascl:1407.007
Shcherbakov, R. V., & Huang, L. 2011, MNRAS, 410, 1052
Shcherbakov, R. V., Penna, R. F., & McKinney, J. C. 2012, ApJ, 755, 133
Shiokawa, H., Gammie, C. F., & Doeleman, S. S. 2017, ApJ, 846, 29
Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 2017, Interferometry and Synthesis in Radio Astronomy (3rd ed.; Cham: Springer)
Twiss, R. Q., Carter, A. W. L., & Little, A. G. 1960, Obs, 80, 153
Twiss, R. Q., Carter, A. W. L., & Little, A. G. 1962, AuJPh, 15, 378
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2019, arXiv:1903.08008
Vousden, W. D., Farr, W. M., & Mandel, I. 2016, MNRAS, 455, 1919
Wardle, J. F. C. 1971, ApL, 8, 53
Younsi, Z., Wu, K., & Fuerst, S. V. 2012, A&A, 545, A13
Younsi, Z., Zhidenko, A., Rezzolla, L., Konoplya, R., & Mizuno, Y. 2016, PhRvD, 94, 084025
Yuan, F., Markoff, S., & Falcke, H. 2002, A&A, 383, 854
Yuan, F., & Narayan, R. 2014, ARA&A, 52, 529
Yuan, F., Quataert, E., & Narayan, R. 2004, ApJ, 606, 894
Zhao, J.-H., Young, K. H., Herrnstein, R. M., et al. 2003, ApJL, 586, L29