Draft version February 4, 2021
Typeset using LaTeX **twocolumn** style in AASTeX62

First M87 Event Horizon Telescope Results VII: polarization of the ring

The Event Horizon Telescope Collaboration

Kazunori Akiyama,[1,2,3,4] Juan Carlos Algaba,[5] Antxon Alberdi,[6] Walter Alef,[7] Richard Anantua,[3,8,9] Keiichi Asada,[10] Rebecca Azulay,[11,12,7] Anne-Kathrin Baczko,[7] David Ball,[13] Mislav Baloković,[14,15] John Barrett,[1] Dan Bintley,[16] Lindy Blackburn,[3,8] Raymond Blundell,[8] Wilfred Boland,[17] Katherine L. Bouman,[3,8,18] Geoffrey C. Bower,[19] Hope Boyce,[20,21] Michael Bremer,[22] Christiaan D. Brinkerink,[23] Roger Brissenden,[3,8] Silke Britzen,[7] Avery E. Broderick,[24,25,26] Dominique Broguiere,[22] Thomas Bronzwaer,[23] Do-Young Byun,[27,28] John E. Carlstrom,[29,30,31,32] Andrew Chael,[33,34] Chi-kwan Chan,[13,35] Shami Chatterjee,[36] Koushik Chatterjee,[37] Ming-Tang Chen,[19] Yongjun Chen (陈永军),[38,39] Paul M. Chesler,[3] Ilje Cho,[27,28] Pierre Christian,[13] John E. Conway,[40] James M. Cordes,[36] Geoffrey B. Crew,[1] Alejandro Cruz-Osorio,[41] Yuzhu Cui,[42,43] Jordy Davelaar,[44,9,23] Mariafelicia De Laurentis,[45,41,46] Roger Deane,[47,48,49] Jessica Dempsey,[16] Gregory Desvignes,[50] Jason Dexter,[51] Sheperd S. Doeleman,[3,8] Ralph P. Eatough,[52,7] Heino Falcke,[23] Joseph Farah,[8,3,53] Vincent L. Fish,[1] Ed Fomalont,[4] Raquel Fraga-Encinas,[23] William T. Freeman,[54,55] Per Friberg,[16] Christian M. Fromm,[3,8,41] Antonio Fuentes,[6] Peter Galison,[3,56,57] Charles F. Gammie,[58,59] Roberto García,[22] Olivier Gentaz,[22] Boris Georgiev,[25,26] Ciriaco Goddi,[23,60] Roman Gold,[61,24] José L. Gómez,[6] Arturo I. Gómez-Ruiz,[62,63] Minfeng Gu (顾敏峰),[38,64] Mark Gurwell,[8] Kazuhiro Hada,[42,43] Daryl Haggard,[20,21] Michael H. Hecht,[1] Ronald Hesper,[65] Luis C. Ho (何子山),[66,67] Paul Ho,[10] Mareki Honma,[42,43,68] Chih-Wei L. Huang,[10] Lei Huang (黄磊),[38,64] David H. Hughes,[62] Shiro Ikeda,[2,69,70,71] Makoto Inoue,[10] Sara Issaoun,[23] David J. James,[3,8] Buell T. Jannuzi,[13] Michael Janssen,[7] Britton Jeter,[25,26] Wu Jiang (江悟),[38] Alejandra Jimenez-Rosales,[23] Michael D. Johnson,[3,8] Svetlana Jorstad,[72,73] Taehyun Jung,[27,28] Mansour Karami,[24,25] Ramesh Karuppusamy,[7] Tomohisa Kawashima,[2] Garrett K. Keating,[8] Mark Kettenis,[74] Dong-Jin Kim,[7] Jae-Young Kim,[7] Jongsoo Kim,[27] Junhan Kim,[13,18] Motoki Kino,[2,75] Jun Yi Koay,[10] Yutaro Kofuji,[42,68] Patrick M. Koch,[10] Shoko Koyama,[10] Michael Kramer,[7] Carsten Kramer,[22] Thomas P. Krichbaum,[7] Cheng-Yu Kuo,[76] Tod R. Lauer,[77] Sang-Sung Lee,[27] Aviad Levis,[18] Yan-Rong Li (李彦荣),[78] Zhiyuan Li (李志远),[79,80] Michael Lindqvist,[40] Rocco Lico,[6,7] Greg Lindahl,[8] Jun Liu (刘俊),[7] Kuo Liu,[7] Elisabetta Liuzzo,[81] Wen-Ping Lo,[10,82] Andrei P. Lobanov,[7] Laurent Loinard,[83,84] Colin Lonsdale,[1] Ru-Sen Lu (路如森),[38,39,7] Nicholas R. MacDonald,[7] Jirong Mao (毛基荣),[85,86,87] Nicola Marchili,[81,7] Sera Markoff,[37,88] Daniel P. Marrone,[13] Alan P. Marscher,[72] Iván Martí-Vidal,[11,12] Satoki Matsushita,[10] Lynn D. Matthews,[1] Lia Medeiros,[89,13] Karl M. Menten,[7] Izumi Mizuno,[16] Yosuke Mizuno,[90,41] James M. Moran,[3,8] Kotaro Moriyama,[1,42] Monika Moscibrodzka,[23] Cornelia Müller,[7,23] Gibwa Musoke,[37,23] Alejandro Mus Mejías,[11,12] Hiroshi Nagai,[2,43] Neil M. Nagar,[91] Masanori Nakamura,[92,10] Ramesh Narayan,[3,8] Gopal Narayanan,[93] Iniyan Natarajan,[49,47,94] Antonios Nathanail,[41,95] Joseph Neilsen,[96] Roberto Neri,[22] Chunchong Ni,[25,26] Aristeidis Noutsos,[7] Michael A. Nowak,[97] Hiroki Okino,[42,68] Héctor Olivares,[23] Gisela N. Ortiz-León,[7] Tomoaki Oyama,[42] Feryal Özel,[13] Daniel C. M. Palumbo,[3,8] Jongho Park,[10] Nimesh Patel,[8] Ue-Li Pen,[24,98,99,100] Dominic W. Pesce,[3,8] Vincent Piétu,[22] Richard Plambeck,[101] Aleksandar PopStefanija,[93] Oliver Porth,[37,41] Felix M. Pötzl,[7] Ben Prather,[58] Jorge A. Preciado-López,[24] Dimitrios Psaltis,[13] Hung-Yi Pu,[102,10,24] Venkatessh Ramakrishnan,[91] Ramprasad Rao,[19] Mark G. Rawlings,[16] Alexander W. Raymond,[3,8] Luciano Rezzolla,[41] Angelo Ricarte,[3,8] Bart Ripperda,[103,9] Freek Roelofs,[23] Alan Rogers,[1] Eduardo Ros,[7] Mel Rose,[13] Arash Roshanineshat,[13] Helge Rottmann,[7] Alan L. Roy,[7] Chet Ruszczyk,[1] Kazi L. J. Rygl,[81] Salvador Sánchez,[104] David Sánchez-Arguelles,[62,105] Mahito Sasada,[42,106] Tuomas Savolainen,[107,108,7] F. Peter Schloerb,[93] Karl-Friedrich Schuster,[22] Lijing Shao,[7,67] Zhiqiang Shen (沈志强),[38,39] Des Small,[74] Bong Won Sohn,[27,28,109] Jason SooHoo,[1] He Sun (孙赫),[18] Fumie Tazaki,[42] Paul Tiede,[25,26] Remo P. J. Tilanus,[23,60,110,13] Michael Titus,[1] Kenji Toma,[111,112] Pablo Torne,[7,104] Tyler Trent,[13] Efthalia Traianou,[7] Sascha Trippe,[113] Ilse van Bemmel,[74] Huib Jan van Langevelde,[74,114] Daniel R. van Rossum,[23] Jan Wagner,[7] Derek Ward-Thompson,[115] John Wardle,[116] Jonathan Weintroub,[3,8] Norbert Wex,[7] Robert Wharton,[7] Maciek Wielgus,[3,8] George N. Wong,[58] Qingwen Wu (吴庆文),[117] Doosoo Yoon,[37] André Young,[23] Ken Young,[8] Ziri Younsi,[118,41] Feng Yuan (袁峰),[38,64,119] Ye-Fei Yuan (袁业飞),[120] J. Anton Zensus,[7] Guang-Yao Zhao,[6] and Shan-Shan Zhao[38]

[1] *Massachusetts Institute of Technology Haystack Observatory, 99 Millstone Road, Westford, MA 01886, USA*

[2] *National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

[3] *Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA*

[4] *National Radio Astronomy Observatory, 520 Edgemont Rd, Charlottesville, VA 22903, USA*

[5] *Department of Physics, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

[6] *Instituto de Astrofísica de Andalucía-CSIC, Glorieta de la Astronomía s/n, E-18008 Granada, Spain*

[7] *Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*

[8]Center for Astrophysics — Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[9]Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[10]Institute of Astronomy and Astrophysics, Academia Sinica, 11F of Astronomy-Mathematics Building, AS/NTU No. 1, Sec. 4, Roosevelt Rd, Taipei 10617, Taiwan, R.O.C.
[11]Departament d'Astronomia i Astrofísica, Universitat de València, C. Dr. Moliner 50, E-46100 Burjassot, València, Spain
[12]Observatori Astronòmic, Universitat de València, C. Catedrático José Beltrán 2, E-46980 Paterna, València, Spain
[13]Steward Observatory and Department of Astronomy, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA
[14]Yale Center for Astronomy & Astrophysics, 52 Hillhouse Avenue, New Haven, CT 06511, USA
[15]Department of Physics, Yale University, P.O. Box 2018120, New Haven, CT 06520, USA
[16]East Asian Observatory, 660 N. A'ohoku Place, Hilo, HI 96720, USA
[17]Nederlandse Onderzoekschool voor Astronomie (NOVA), PO Box 9513, 2300 RA Leiden, The Netherlands
[18]California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA
[19]Institute of Astronomy and Astrophysics, Academia Sinica, 645 N. A'ohoku Place, Hilo, HI 96720, USA
[20]Department of Physics, McGill University, 3600 rue University, Montréal, QC H3A 2T8, Canada
[21]McGill Space Institute, McGill University, 3550 rue University, Montréal, QC H3A 2A7, Canada
[22]Institut de Radioastronomie Millimétrique, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France
[23]Department of Astrophysics, Institute for Mathematics, Astrophysics and Particle Physics (IMAPP), Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[24]Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, N2L 2Y5, Canada
[25]Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
[26]Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1 Canada
[27]Korea Astronomy and Space Science Institute, Daedeok-daero 776, Yuseong-gu, Daejeon 34055, Republic of Korea
[28]University of Science and Technology, Gajeong-ro 217, Yuseong-gu, Daejeon 34113, Republic of Korea
[29]Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[30]Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[31]Department of Physics, University of Chicago, 5720 South Ellis Avenue, Chicago, IL 60637, USA
[32]Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[33]Princeton Center for Theoretical Science, Jadwin Hall, Princeton University, Princeton, NJ 08544, USA
[34]NASA Hubble Fellowship Program, Einstein Fellow
[35]Data Science Institute, University of Arizona, 1230 N. Cherry Ave., Tucson, AZ 85721, USA
[36]Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA
[37]Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
[38]Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, People's Republic of China
[39]Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China
[40]Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, SE-43992 Onsala, Sweden
[41]Institut für Theoretische Physik, Goethe-Universität Frankfurt, Max-von-Laue-Straße 1, D-60438 Frankfurt am Main, Germany
[42]Mizusawa VLBI Observatory, National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu, Iwate 023-0861, Japan
[43]Department of Astronomical Science, The Graduate University for Advanced Studies (SOKENDAI), 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[44]Department of Astronomy and Columbia Astrophysics Laboratory, Columbia University, 550 W 120th Street, New York, NY 10027, USA
[45]Dipartimento di Fisica "E. Pancini", Universitá di Napoli "Federico II", Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
[46]INFN Sez. di Napoli, Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
[47]Wits Centre for Astrophysics, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, Johannesburg 2050, South Africa
[48]Department of Physics, University of Pretoria, Hatfield, Pretoria 0028, South Africa
[49]Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, Makhanda 6140, South Africa
[50]LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France
[51]JILA and Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA
[52]National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100101, PR China
[53]University of Massachusetts Boston, 100 William T. Morrissey Boulevard, Boston, MA 02125, USA
[54]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 32-D476, 77 Massachusetts Ave., Cambridge, MA 02142, USA
[55]Google Research, 355 Main St., Cambridge, MA 02142, USA
[56]Department of History of Science, Harvard University, Cambridge, MA 02138, USA

[57]Department of Physics, Harvard University, Cambridge, MA 02138, USA
[58]Department of Physics, University of Illinois, 1110 West Green Street, Urbana, IL 61801, USA
[59]Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA
[60]Leiden Observatory—Allegro, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[61]CP3-Origins, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
[62]Instituto Nacional de Astrofísica, Óptica y Electrónica. Apartado Postal 51 y 216, 72000. Puebla Pue., México
[63]Consejo Nacional de Ciencia y Tecnologıa, Av. Insurgentes Sur 1582, 03940, Ciudad de México, México
[64]Key Laboratory for Research in Galaxies and Cosmology, Chinese Academy of Sciences, Shanghai 200030, People's Republic of China
[65]NOVA Sub-mm Instrumentation Group, Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
[66]Department of Astronomy, School of Physics, Peking University, Beijing 100871, People's Republic of China
[67]Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
[68]Department of Astronomy, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[69]The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8562, Japan
[70]Department of Statistical Science, The Graduate University for Advanced Studies (SOKENDAI), 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
[71]Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, 277-8583, Japan
[72]Institute for Astrophysical Research, Boston University, 725 Commonwealth Ave., Boston, MA 02215, USA
[73]Astronomical Institute, St. Petersburg University, Universitetskij pr., 28, Petrodvorets,198504 St.Petersburg, Russia
[74]Joint Institute for VLBI ERIC (JIVE), Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands
[75]Kogakuin University of Technology & Engineering, Academic Support Center, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan
[76]Physics Department, National Sun Yat-Sen University, No. 70, Lien-Hai Rd, Kaosiung City 80424, Taiwan, R.O.C
[77]National Optical Astronomy Observatory, 950 North Cherry Ave., Tucson, AZ 85719, USA
[78]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
[79]School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China
[80]Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, Nanjing 210023, People's Republic of China
[81]Italian ALMA Regional Centre, INAF-Istituto di Radioastronomia, Via P. Gobetti 101, I-40129 Bologna, Italy
[82]Department of Physics, National Taiwan University, No.1, Sect.4, Roosevelt Rd., Taipei 10617, Taiwan, R.O.C
[83]Instituto de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, Morelia 58089, México
[84]Instituto de Astronomía, Universidad Nacional Autónoma de México, CdMx 04510, México
[85]Yunnan Observatories, Chinese Academy of Sciences, 650011 Kunming, Yunnan Province, People's Republic of China
[86]Center for Astronomical Mega-Science, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing, 100012, People's Republic of China
[87]Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, 650011 Kunming, People's Republic of China
[88]Gravitation Astroparticle Physics Amsterdam (GRAPPA) Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[89]School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA
[90]Tsung-Dao Lee Institute and School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, China
[91]Astronomy Department, Universidad de Concepción, Casilla 160-C, Concepción, Chile
[92]National Institute of Technology, Hachinohe College, 16-1 Uwanotai, Tamonoki, Hachinohe City, Aomori 039-1192, Japan
[93]Department of Astronomy, University of Massachusetts, 01003, Amherst, MA, USA
[94]South African Radio Astronomy Observatory, Observatory 7925, Cape Town, South Africa
[95]Department of Physics, National and Kapodistrian University of Athens, Panepistimiopolis, GR 15783 Zografos, Greece
[96]Villanova University, Mendel Science Center Rm. 263B, 800 E Lancaster Ave, Villanova PA 19085
[97]Physics Department, Washington University CB 1105, St Louis, MO 63130, USA
[98]Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada
[99]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[100]Canadian Institute for Advanced Research, 180 Dundas St West, Toronto, ON M5G 1Z8, Canada
[101]Radio Astronomy Laboratory, University of California, Berkeley, CA 94720, USA
[102]Department of Physics, National Taiwan Normal University, No. 88, Sec.4, Tingzhou Rd., Taipei 116, Taiwan, R.O.C.
[103]Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
[104]Instituto de Radioastronomía Milimétrica, IRAM, Avenida Divina Pastora 7, Local 20, E-18012, Granada, Spain
[105]Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, 03940, Ciudad de México, México
[106]Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
[107]Aalto University Department of Electronics and Nanoengineering, PL 15500, FI-00076 Aalto, Finland

[108] *Aalto University Metsähovi Radio Observatory, Metsähovintie 114, FI-02540 Kylmälä, Finland*
[109] *Department of Astronomy, Yonsei University, Yonsei-ro 50, Seodaemun-gu, 03722 Seoul, Republic of Korea*
[110] *Netherlands Organisation for Scientific Research (NWO), Postbus 93138, 2509 AC Den Haag, The Netherlands*
[111] *Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan*
[112] *Astronomical Institute, Tohoku University, Sendai 980-8578, Japan*
[113] *Department of Physics and Astronomy, Seoul National University, Gwanak-gu, Seoul 08826, Republic of Korea*
[114] *Leiden Observatory, Leiden University, Postbus 2300, 9513 RA Leiden, The Netherlands*
[115] *Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK*
[116] *Physics Department, Brandeis University, 415 South Street, Waltham, MA 02453, USA*
[117] *School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, People's Republic of China*
[118] *Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK*
[119] *School of Astronomy and Space Sciences, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, People's Republic of China*
[120] *Astronomy Department, University of Science and Technology of China, Hefei 230026, People's Republic of China*

ABSTRACT

In April 2017, the Event Horizon Telescope observed the near-horizon region around the supermassive black hole at the core of the M87 galaxy. These 1.3 mm wavelength observations revealed a compact asymmetric ring-like source morphology. Such a structure originates from synchrotron emission produced by relativistic plasma located in the immediate vicinity of the black hole. Here we present the corresponding linear-polarimetric images of the center of M87. We find that only a part of the ring is significantly polarized. The resolved fractional linear polarization has a maximum located in the south-west part of the ring, where it rises to the level of $\sim$15 %. The polarization position angles are arranged into a nearly azimuthal pattern. We perform quantitative measurements of relevant polarimetric properties of the compact emission and find evidence for the temporal evolution of the polarized source structure over one week of EHT observations. The details of the polarimetric data reduction and calibration methodology are provided. We carry out the data analysis using multiple independent imaging and modeling techniques, each of which is validated against a suite of synthetic datasets. The gross polarimetric structure and its apparent evolution with time are insensitive to the method used to reconstruct the image. These polarimetric images carry information about the structure of the magnetic fields responsible for the synchrotron emission. Their physical interpretation is discussed in an accompanying publication.

Keywords: Galaxies: individual: M87; Radio interferometry; Very long baseline interferometry; Polarimetry; Supermassive black holes; Active galactic nuclei; Low-luminosity active galactic nuclei; Astronomy data modeling; Galaxy accretion disks

Contents

1. INTRODUCTION

The Event Horizon Telescope (EHT) collaboration has recently reported the first images of the event-horizon scale structure around the supermassive black hole in the core of the massive elliptical galaxy M 87 – one of its two main targets[1]. The EHT images of M87's core at 230 GHz revealed a ring-like structure whose diameter of 42 μas, brightness temperature, shape and asymmetry are interpreted as the black hole's "shadow" illuminated by synchrotron emission from relativistic electrons gyrating around magnetic field lines in close vicinity to the event horizon. We have described the details of the EHT's instrumentation, data calibration pipelines, data analyses and imaging procedures, and the theoretical interpretation of these first images in a series of publications (Event Horizon Telescope Collaboration et al. 2019a,b,c,d,e,f, hereafter Papers I–VI).

In this paper, we present the first *polarimetric* analysis of the EHT 2017 observations of M87 and the first images of the linearly polarized radiation surrounding the M87 black hole shadow. These polarimetric images provide essential new information about the structure of magnetic field lines near the event horizon of M87's central supermassive black hole, and they put tight constraints on the theoretical interpretations of the nature of the ring and of relativistic jet launching theories. The theoretical implications of these images and the constraints they place on the magnetic field structure and accretion state of the black hole are discussed in an accompanying work (Event Horizon Telescope Collaboration et al. 2020, hereafter Paper VIII). Readers interested in the details of the data reduction, methodology, and validation can find a detailed index of the paper in Section 1.2. Readers primarily interested in the results may skip directly to Section 5 and the subsequent discussion and conclusions in Sections 6.

1.1. *Previous polarimetric observations of M87 jet*

The giant elliptical galaxy Messier 87 (M87, NGC 4486) is the central member of the Virgo cluster of galaxies and hosts a low luminosity radio source (Virgo A, 3C 274, B1226+126). M87 is nearby and bright, and at its center is one of the best studied Active Galactic Nuclei (AGN). M87 was the first galaxy in which an extragalactic jet (first described as a "narrow ray") extending from the nucleus was discovered (Curtis 1918). This kiloparsec scale jet is visible, with remarkably similar morphology, at all wavelengths from radio to X-ray. The optical radiation from the jet on kpc scales was found to be linearly polarized by Baade (1956), which was confirmed by Hiltner (1959), suggesting that the emission mechanism is synchrotron radiation.

The central engine that powers the jet contains one of the most massive black holes known, measured from the central stellar velocity dispersion (Gebhardt et al. 2011; $M = (6.6 \pm 0.4) \times 10^9 M_\odot$) and directly from the size of the observed emitting region surrounding the black

[1] The other primary target being the black hole Sgr A* in the center of the Milky Way.

hole shadow (Paper VI; $M = (6.5 \pm 0.7) \times 10^9 \mathrm{M}_\odot$). For this mass, the Schwarzschild radius is $R_\mathrm{s} = 2GM/c^2 = 1.8 \times 10^{15}$ cm. At the distance of M87, $16.8^{+0.8}_{-0.7}$ Mpc (Blakeslee et al. 2009, Bird et al. 2010, Cantiello et al. 2018, Paper VI), the EHT resolution of about 20 microarcseconds (μas) translates into linear scale of 0.0016 pc $= 2.5\,R_\mathrm{s}$.

The M87 jet has been imaged at sub-arcsecond resolution in both total intensity and linear polarization at optical wavelengths with the *Hubble Space Telescope* (Thomson et al. 1995; Capetti et al. 1997), and at radio wavelengths with the Very Large Array (e.g., Owen et al. 1989). Observing the launching region of the jet closer to the black hole and the region surrounding the black hole requires milli-arcsecond (mas) resolution or better, and hence very-long-baseline interferometry (VLBI) techniques used at the highest frequencies (e.g., Boccardi et al. 2017 and references therein).

Milli-arcsecond-scale VLBI observations show that the core itself is unpolarized even at millimeter wavelengths. Zavala & Taylor (2002), observing at 8, 12, and 15 GHz, set upper limits on the fractional polarization of the compact core of $m < 0.1\%$. About 20 mas downstream from the core, patchy linear polarization starts to become visible in the jet at the level of $5-10\%$, although no large-scale coherent pattern to the electric vector position angles (EVPAs) χ is apparent. However, at each patch in the downstream jet, the EVPAs exhibit a linear change with λ^2, allowing the Rotation Measures (RMs) to be estimated. These RMs range from $-4000\,\mathrm{rad\,m^{-2}}$ to $9000\,\mathrm{rad\,m^{-2}}$ (Zavala & Taylor 2002). The linear dependence of EVPA on λ^2 over several radians is important, as it shows that the Faraday-rotating plasma in the jet cannot be mixed in with the relativistic emitting particles (Burn 1966) but must be in a cooler (sub-relativistic) foreground screen. On kiloparsec scales, Owen et al. (1990) find a complex distribution of rotation measure. Over most of the source the rotation measure is typically of order $1000\,\mathrm{rad\,m^{-2}}$, but there are patches where values as high as $8000\,\mathrm{rad\,m^{-2}}$ are found.

More recently, Park et al. (2019) studied Faraday rotation measures in the jet using multifrequency VLBA data at $\lesssim 8$ GHz. They found that the magnitude of RM systematically decreases with increasing distance from 5,000 to 200,000 Schwarzschild radii. The observed large ($\gtrsim 45$ deg) EVPA rotations at various locations of the jet suggest that the dominant Faraday screen in this distance range would be external to the jet, similar to the conclusion of Zavala & Taylor (2002). Homan & Lister (2006), also observing at 15 GHz with the Very Long Baseline Array (VLBA; as part of the MOJAVE project) found a tight upper limit on the fractional linear polarization of the core of $< 0.07\%$. They also detect circular polarization of $(-0.49 \pm 0.10)\%$.

At 43 GHz, Walker et al. (2018) presented results from 17 years of VLBA observations of M87, with polarimetric images presented at two epochs. These show significant polarization (up to 4%) in the jet near the 43 GHz core, but at the position of the total intensity peaks the fractional polarizations are only 1.5% and 1.1%. They interpret these fractions as coming from a mix of emission from the unresolved, unpolarized core and a more polarized inner jet.

Hada et al. (2016) show images at four epochs at 86 GHz made with the VLBA and the Green Bank Telescope. At this frequency, the resolution is about (0.4×0.1) mas, corresponding to $(56 \times 14)\,R_s$. Again, the core is unpolarized with no linear polarization detected at the position of the total intensity emission's peak, while there is a small patch of significant (3.5%) polarization located 0.1 mas downstream. At 0.4 mas downstream from the peak, there is another patch of significant polarization (20%). These results indicate that there are regions of significantly ordered magnetic field very close to the central engine.

Very recently, new observations by Kravchenko et al. (2020) using the VLBA at 22 and 43 GHz show two components of linear polarization and a smooth rotation of EVPA around the 43 GHz core. Comparison with earlier observations show that the global polarization pattern in the jet is largely stable over an 11 year timescale. They suggest that the polarization pattern is associated with the magnetic structure in a confining magnetohydrodynamic wind, which is also the source of the observed Faraday rotation.

The EHT presently observes at ∼230 GHz and has previously reported polarimetric measurements only for Sgr A* (Johnson et al. 2015). The only previous polarimetric measurements of M87 at this frequency are done by Kuo et al. (2014) using the Submillimeter Array (SMA) on Mauna Kea, Hawai'i (USA). The SMA is a compact array with a (1.2×0.8) arcsec beam, ten thousand times larger than the EHT beam. Li et al. (2016) use the value from this work to calculate a limit on the accretion rate onto the M87 black hole. Most recently, Goddi et al in prep (2020) reported results on M87 around 230 GHz as part of the ALMA's interferometric connected-element array portion of the EHT observations in 2017. The ALMA-only 230 GHz observations (with a FWHM synthesized beam in the range 1-2 arcsec, depending on the day) resolve M87 inner region into a compact central core and a kpc-scale jet across approximately 25 arcsec. It has been found that the 230 GHz core at these scales has a total flux of ∼1.3 Jy, a low linear polarization fraction $|m| \sim 2.7\%$, and even less circular polarization, $|v| < 0.3\%$. Notably, ALMA-only observations show strong variability in the rotation measure estimated based on four frequencies within ALMA Band 6 (four spectral windows centered at 213, 215, 227 and 229 GHz, Matthews et al. 2018). The RM difference is clear between the start of the EHT observing campaign on April 5 (RM $\approx 0.6 \times 10^5$ rad m^{-2}) and the end on April 11 (RM $\approx -0.4 \times 10^5$ rad m^{-2}). Since these measurements were taken simultaneously with the

EHT VLBI observations presented here, the ALMA-only linear polarization fraction measurements can be used as a point of reference, and we discuss possible implications of the strong RM evolution on the EHT polarimetric images of M87.

1.2. *This work*

This paper presents the details of the polarimetric data calibration, the procedures for polarimetric imaging, and the resulting images of the M87 core. In Section 2, we briefly overview the basics of polarimetric VLBI. In Section 3, we summarize the EHT 2017 observations, describe the initial data calibration procedure and validation tests, and we describe the basic properties of the polarimetric data. In Section 4, we describe our methods, strategy, test suite for our polarimetric leakage calibration and polarimetric imaging. In Section 5, we present and analyse the polarimetric images of the M87 ring and we examine the calibration's impact on the polarimetric image. We discuss the results and summarize the work in Sections 6 and 7.

The paper is supplemented with a number of appendices supporting our analysis and results. The appendices summarize: polarimetric data issues (Appendix A); novel VLBI closure data products (Appendix B); details of calibration and imaging methods (Appendix C); validation of polarimetric calibration for telescopes with an intra-site partner (Appendix D); fiducial leakage D-terms from M87 imaging (Appendix E); preliminary results of polarimetric imaging of M87 (Appendix F); polarimetric imaging scoring procedures (Appendix G); details of Monte Carlo D-term simulations (Appendix H); consistency of low and high band results for M87 (Appendix I); comparison polarimetric properties of calibrator sources (Appendix J); and validations of assumptions made in polarimetric imaging of the main target and the calibrators (Appendix K).

2. BASIC DEFINITIONS

A detailed introduction to polarimetric VLBI can be found in Thompson et al. (2017), Chapter 4. Here we briefly introduce the basic concepts and notation necessary to understand the analysis presented throughout this paper. The polarized state of the electromagnetic radiation at a given spatial coordinate $\mathbf{x} = (x, y)$ is described in terms of four Stokes parameters, $\mathcal{I}(\mathbf{x})$ (total intensity), $\mathcal{Q}(\mathbf{x})$ (difference in horizontal and vertical linear polarization), $\mathcal{U}(\mathbf{x})$ (difference in linear polarization at 45° and −45° position angle), and $\mathcal{V}(\mathbf{x})$ (circular polarization). We define the complex linear polarization p as

$$p \equiv \mathcal{Q} + i\mathcal{U} = \mathcal{I}|m|e^{2i\chi} \ , \quad (1)$$

where $m = (\mathcal{Q} + i\mathcal{U})/\mathcal{I}$ represents the (complex) fractional polarization, and $\chi = 0.5 \arg(p)$ is the EVPA, measured from North to East. Total intensity VLBI observations directly sample the Fourier transform $\tilde{\mathcal{I}}$ as a function of the spatial frequency $\mathbf{u} = (u, v)$ of the total intensity image; similarly, polarimetric VLBI observations also sample the Fourier transform of the other Stokes parameters $\tilde{\mathcal{Q}}$, $\tilde{\mathcal{U}}$, $\tilde{\mathcal{V}}$.

EHT data are represented in a circular basis, related to the Stokes visibility components with the following coordinate system transformation

$$\rho_{jk} \equiv \begin{pmatrix} R_jR_k^* & R_jL_k^* \\ L_jR_k^* & L_jL_k^* \end{pmatrix} = \begin{pmatrix} \tilde{\mathcal{I}}_{jk} + \tilde{\mathcal{V}}_{jk} & \tilde{\mathcal{Q}}_{jk} + i\tilde{\mathcal{U}}_{jk} \\ \tilde{\mathcal{Q}}_{jk} - i\tilde{\mathcal{U}}_{jk} & \tilde{\mathcal{I}}_{jk} - \tilde{\mathcal{V}}_{jk} \end{pmatrix} \quad (2)$$

for a baseline between two stations j and k. Notation $R_jL_k^*$ indicates complex correlation (where the asterisk denotes conjugation) of the electric field components measured by the telescopes, in this example the right-hand circularly polarized component E_R measured by the telescope j and the left-hand circularly polarized component E_L measured by the telescope k. Equation 2 defines the coherency matrix ρ_{jk}. Following Johnson et al. (2015), we also define the fractional polarization in the visibility domain,

$$\breve{m} \equiv \frac{\tilde{\mathcal{Q}} + i\tilde{\mathcal{U}}}{\tilde{\mathcal{I}}} = \frac{\tilde{p}}{\tilde{\mathcal{I}}} = \frac{2RL^*}{RR^* + LL^*} \ . \quad (3)$$

Note that Equation 3 implies that $\breve{m}(\mathbf{u})$ and $\breve{m}(-\mathbf{u})$ constitute independent measurements for $\mathbf{u} \neq 0$. Moreover, $\breve{m}(\mathbf{u})$ and $m(\mathbf{x})$ are *not* a Fourier pair. While the image domain fractional polarization magnitude is restricted to values between 0 (unpolarized radiation) and 1 (full linear polarization), there is no such restriction on the absolute value of $\breve{m}$. Useful relationships between $\breve{m}$ and m are discussed in Johnson et al. (2015).

Imperfections in the instrumental response distort the relationship between the measured polarimetric visibilities and the source's intrinsic polarization. These imperfections can be conveniently described by a Jones matrix formalism (Jones 1941), and estimates of the Jones matrix coefficients can then be used to correct the distortions. The Jones matrix characterizing a particular station can be decomposed into a series of complex matrices $\mathbf{G}$, $\mathbf{D}$ and $\mathbf{\Phi}$ (Thompson et al. 2017),

$$\mathbf{J} = \mathbf{GD\Phi} = \begin{pmatrix} G_R & 0 \\ 0 & G_L \end{pmatrix} \begin{pmatrix} 1 & D_R \\ D_L & 1 \end{pmatrix} \begin{pmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix} . \quad (4)$$

Time-dependent field rotation matrices $\mathbf{\Phi} \equiv \mathbf{\Phi}(t)$ are known a priori, with the field rotation angle $\phi(t)$ dependent on the source's elevation $\theta_{\rm el}(t)$ and parallactic angle $\psi_{\rm par}(t)$. The angle ϕ takes the general form of

$$\phi = f_{\rm el}\theta_{\rm el} + f_{\rm par}\psi_{\rm par} + \phi_{\rm off} \ , \quad (5)$$

where $\phi_{\rm off}$ is a constant offset. The coefficients $f_{\rm el}$ and $f_{\rm par}$ are specific to the receiver position type. The gain matrices $\mathbf{G}$, containing complex station gains G_R and G_L, are estimated within the EHT's upstream calibration and total intensity imaging pipeline, see Section 3.2.

Estimation of the D-terms, the complex coefficients D_R and D_L of the leakage matrix $\mathbf{D}$, generally requires simultaneous modeling of the resolved calibration source, and hence cannot be easily applied at the upstream data calibration stage. The details of the leakage calibration procedures adopted for the EHT polarimetric data sets analysis are described in Section 4.

For a pair of VLBI stations j and k the measured coherency matrix ρ'_{jk} is related to the true source coherency matrix ρ_{jk} via the Radio Interferometer Measurement Equation, hereafter RIME (Hamaker et al. 1996; Smirnov 2011),

$$\rho'_{jk} = \mathbf{J}_j \rho_{jk} \mathbf{J}_k^\dagger \,, \tag{6}$$

where the dagger † symbol denotes conjugate transposition. Once the Jones matrices for the stations j and k are well characterized, Equation 6 can be inverted to give the source coherency matrix ρ_{jk}. From ρ_{jk}, Stokes visibilities can be obtained by inverting Equation 2:

$$\begin{pmatrix} \tilde{\mathcal{I}}_{jk} \\ \tilde{\mathcal{Q}}_{jk} \\ \tilde{\mathcal{U}}_{jk} \\ \tilde{\mathcal{V}}_{jk} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} R_j R_k^* + L_j L_k^* \\ R_j L_k^* + L_j R_k^* \\ -i(R_j L_k^* - L_j R_k^*) \\ R_j R_k^* - L_j L_k^* \end{pmatrix} . \tag{7}$$

The collection of Stokes visibilities sampled in (u, v) space by the VLBI array can finally be used to reconstruct the polarimetric images $\mathcal{I}(\mathbf{x})$, $\mathcal{Q}(\mathbf{x})$,$\mathcal{U}(\mathbf{x})$, and $\mathcal{V}(\mathbf{x})$.

The coherency matrices on a quadrangle of baselines can be combined to form "closure traces," data products that are insensitive to any calibration effects that can be described using Jones matrices. Appendix B defines these closure traces and outlines their utility for describing the EHT data.

3. EHT 2017 POLARIMETRIC DATA

3.1. *Observations and Initial Processing*

Eight observatories at six geographical locations participated in the 2017 EHT observing campaign: the Atacama Large Millimeter/submillimeter Array (ALMA) and the Atacama Pathfinder Experiment (APEX) in the Atacama Desert in Chile; the Large Millimeter Telescope Alfonso Serrano (LMT) on the Volcán Sierra Negra in Mexico; the South Pole Telescope (SPT) at the geographic south pole; the IRAM 30 m telescope (PV) on Pico Veleta in Spain; the Submillimeter Telescope (SMT) on Mt. Graham in Arizona (USA); the Submillimeter Array (SMA) and the James Clerk Maxwell Telescope (JCMT) on Maunakea in Hawaiʻi (USA) [2].

[2] In the EHT array, there are stations with a co-located element of the array: ALMA and APEX (with ∼2 km baseline) and JCMT and SMA (with ∼0.2 km baseline). We further refer to these two baselines as *zero* baselines or *intra-site* baselines.

The EHT observations were carried out on five nights between April 5–11, 2017. M87 was observed on April 5, 6, 10 and 11. Along with the main EHT targets M87 and Sagittarius A*, several other AGN sources were observed as science targets and calibrators.

Observations were conducted using two contiguous frequency bands of 2 GHz bandwidth each, centered at frequencies of 227.1 and 229.1 GHz, hereby referred to as low and high band, respectively. Apart from the JCMT, which observed only a single polarization (right circular polarization on April 5–7 and left circular polarization on April 10–11), all stations observed in full polarization mode. ALMA is the only station to natively record data in a linear polarization basis. Visibilities measured on baselines to ALMA were converted from a mixed linear-circular basis to circular polarization after correlation using the `PolConvert` software (Martí-Vidal et al. 2016; Matthews et al. 2018; Goddi et al. 2019). A technical description of the EHT array is presented in Paper II and a summary of the 2017 observations and data reduction is presented in Paper III.

3.2. *Correlation and Data Calibration*

After the sky signal received at each telescope was mixed to baseband, digitized, and recorded directly to hard disk, the data from each station were sent to MIT Haystack Observatory and the Max-Planck-Institut für Radioastronomie (MPIfR) for correlation using the DiFX software correlators (Deller et al. 2011). The accumulation period adopted at correlation is 0.4s, with a frequency resolution of 0.5 MHz. The clock model used during correlation to align the wavefronts arriving at different telescopes is imperfect, owing to an approximate a priori model for Earth geometry as well as rapid stochastic variations in path length due to local atmospheric turbulence (Paper III). Before the data can be averaged coherently to build up Signal-to-Noise ratio (S/N), these effects must be accurately measured and corrected. This process, referred to as fringe-fitting, was conducted using three independent software packages: the Haystack Observatory Processing System (HOPS; Whitney et al. 2004; Blackburn et al. 2019); the Common Astronomy Software Applications package (CASA; McMullin et al. 2007; Janssen et al. 2019b); and the NRAO Astronomical Image Processing System (AIPS; Greisen 2003, Paper III). Automated reduction pipelines were designed specifically to address the unique challenges related to the heterogeneity, wide bandwidth, and high observing frequency of EHT data. Field rotation angle is corrected with Equations 4-5, using coefficients given in Table 1. Flux density (amplitude) calibration is applied via a common post-processing framework for all pipelines (Blackburn et al. 2019, Paper III), taking into account estimated station sensitivities (Issaoun et al. 2017; Janssen et al. 2019a). Under the assumption of zero circular polarization of the primary (solar system) calibrator sources,

Table 1. Field rotation parameters for the EHT stations.

Station	Receiver location	$f_{\rm par}$	$f_{\rm el}$	$\phi_{\rm off}$(deg)
ALMA	Cassegrain	1	0	0
APEX	Nasmyth-Right	1	1	0
JCMT	Cassegrain	1	0	0
SMA	Nasmyth-Left	1	-1	45
LMT	Nasmyth-Left	1	-1	0
SMT	Nasmyth-Right	1	1	0
PV	Nasmyth-Left	1	-1	0
SPT	Cassegrain	1	0	0

elevation-independent station gains possess independent statistical uncertainties for the RCP and LCP signal paths, estimated to be $\sim 20\%$ for the LMT and $\sim 10\%$ for all other stations (Janssen et al. 2019a).

To remove the instrumental amplitude mismatch between the LL^* and RR^* visibility components (the R-L phases are correctly calibrated in all scans by using ALMA as the reference station), calibration of the complex polarimetric gain ratios (the ratios of the G_R and G_L terms in the $\mathbf{G}$ matrices) is performed. This is done by fitting global (multi-source, multi-days) piecewise polynomial gain ratios as functions of time. The aim of this approach is to preserve differences in LL^* and RR^* visibilities intrinsic to the source (Steel et al. 2019). After this step, preliminary polarimetric Stokes visibilities $\tilde{\mathcal{I}}, \tilde{\mathcal{Q}}, \tilde{\mathcal{U}}, \tilde{\mathcal{V}}$ can be constructed. However, the gain calibration requires significant additional improvements. The final calibration of the station phase and amplitude gains takes place in a self-calibration step as part of imaging or modeling the Stokes $\mathcal{I}$ brightness distribution, preserving the complex polarimetric gain ratios (e.g., Paper IV, Paper VI). Fully calibrating the D-terms requires modeling the polarized emission.

The Stokes $\mathcal{I}$ (total intensity) analysis of a subset of the 2017 observations (Science Release 1; SR1), including M87, was the subject of Paper I–Paper VI. The quality of these Stokes $\mathcal{I}$ data was assured by a series of tests covering self-consistency over bands and parallel hands polarizations, and consistency of trivial closure quantities (Wielgus et al. 2019). Constraints on the residual non-closing errors were found to be at a 2% level.

For additional information on the calibration, data reduction, and validation procedures for EHT, see Paper III. In this paper, we utilize the HOPS pipeline full-polarization band-averaged (i.e., averaged over frequency within each band) and 10-second averaged data set from the same reduction path as SR1, but containing a larger sample of calibrator sources for polarimetric leakage studies. In addition, the ALMA linear-polarization observing mode allows us to measure and recover the absolute EVPA in the calibrated VLBI visibilities (Martí-Vidal et al. 2016). Other minor polarimetric data handlings due to station-specific issues and conventions are presented in Appendix A.

3.3. *Polarimetric Data Properties*

In Figure 1 (top panels) we show the (u, v) coverage and low-band interferometric polarization of our main target M87 as a function of the baseline (u, v) after the initial calibration stage but before D-term calibration. The colors code the scan-averaged amplitude of the complex fractional polarization $\breve{m}$ (i.e., the fractional polarization in visibility space; for analysis of $\breve{m}$ in another source, Sgr A*, see Johnson et al. 2015). M87 is weakly polarized on most baselines, $|\breve{m}| \lesssim 0.5$. Several data points on SMA–SMT baselines have very high fractional polarization $|\breve{m}(u, v)| \sim 2$ that occur at (u, v) spacings where the Stokes $\mathcal{I}$ visibility amplitude enters a deep minimum. The fractional polarization $\breve{m}$ of the M87 core is broadly consistent across the four days of observations and between low and high frequency bands, therefore high-band results are omitted in the display.

In Figure 1 (lower panels) we show the field rotation angles ϕ for each station observing M87 on the four observing days. The data are corrected for this angle during the initial calibration stage, but the precision of the leakage calibration depends on how well this angle is covered and on the difference in the field angles at the two stations forming a baseline. In the M87 data the field rotation for stations forming long baselines (LMT, SMT, and PV) is frequently larger than 100° except for April 10, for which the (u, v) tracks are shorter.

In addition to the M87 data, a number of calibrators are utilized in this paper for leakage calibration studies. To estimate D-terms for each of the EHT stations we use several EHT targets observed near-in-time to M87. In VLBI, weakly polarized sources are more sensitive to polarimetric calibration errors so they are preferred calibrators. For full-array leakage calibration, we focus on two additional sources: J1924–2914 and NRAO 530 (calibrators for the second EHT primary target, Sgr A*), which are compact and relatively weakly polarized. The main calibrator to M87 in total intensity, 3C 279 (Paper IV, Kim et al. 2020), is bright and strongly polarized on longer baselines and is not used in this work. The properties of the calibrators are discussed in more detail in Appendix J.

The closure traces for M87 and the calibrators can be used both to probe the data for uncalibrated systematic effects (see Appendix B.2) and to ascertain the presence of polarized flux density in a calibration-insensitive manner (see Appendix B.3).

Unless otherwise stated the following analysis is focused on the low band half of the data sets.

4. METHODS FOR POLARIMETRIC IMAGING AND LEAKAGE CALIBRATION

4.1. *Methods*

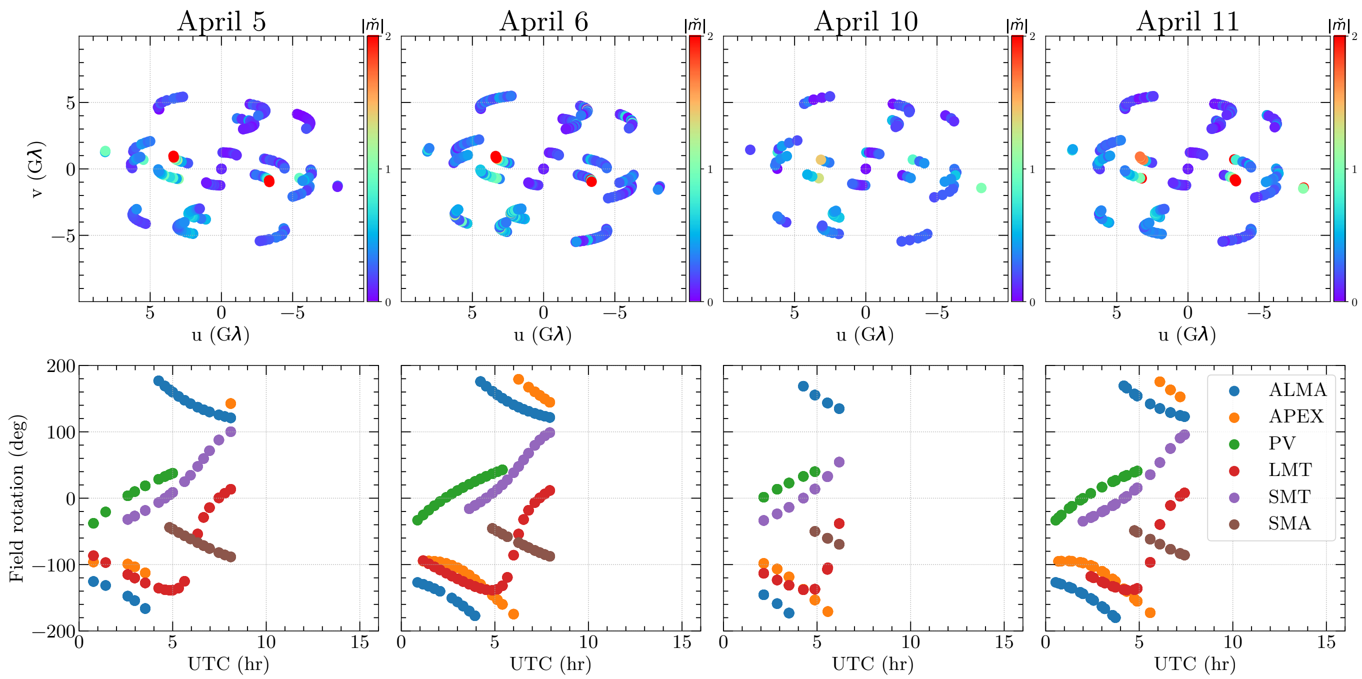

Figure 1. Top: (u, v) coverage of the four M87 observing days in the 2017 campaign. The color of the data points codes the fractional polarization amplitude $|\breve{m}(u, v)|$ in the range from 0 to 2. Error bars on $|\breve{m}|$ are omitted for clarity. A few data points with very high fractional polarization $|\breve{m}(u, v)| \sim 2$ correspond to (u, v) points where the Stokes $\mathcal{I}$ visibility amplitude profile has deep minima (SMA–SMT baselines). The data shown are derived from low-band visibilities after the initial calibration pipeline described in Section 3.2 but before any D-term calibration. The data points are coherently scanned-averaged. Bottom: M87 field rotation angle ϕ for each station as a function of time.

Producing an image of the linearly polarized emission requires both solving for the distribution of Stokes parameters $\mathcal{Q}$ and $\mathcal{U}$ on the sky and for the complex D-terms that mix right and left circular polarization at the stations. In this work, we use several distinct methods to accomplish these tasks. Our approaches can be classified into three main categories: imaging via sub-component fitting; imaging via regularized maximum likelihood; and imaging as posterior exploration. In this section we only briefly describe each method: fuller descriptions are presented in Appendix C.

The calibration of the instrumental polarization by sub-component fitting has been performed using three different codes (`LPCAL`, `GPCAL`, and `polsolve`) that depend on two standard software packages for interferometric data analysis: AIPS[3] and CASA[4]. In all of these methods, the Stokes $\mathcal{I}$ imaging step is performed using the CLEAN algorithm (Högbom 1974), and sub-components with constant complex fractional polarization are constructed from collections of the total intensity CLEAN components and fit to the data. In AIPS, two algorithms for D-term calibration are available: `LPCAL` (extensively used in VLBI polarimetry for more than 20 years; Leppänen et al. 1995) and `GPCAL`[5] (Park et al. 2021). In CASA, we use the `polsolve` algorithm (Martí-Vidal et al., *in prep.*), which uses multiple calibrators simultaneously to fit polarimetric sub-components and allows for frequency-dependent D-terms (see Appendix D). In sub-component fitting and imaging with `LPCAL` and `polsolve` we assume that Stokes $\mathcal{V} = 0$. Further details on `LPCAL`, `GPCAL`, and `polsolve` can be found in Appendix C.1.

Image reconstruction via the Regularized Maximum Likelihood (RML) method was used in Paper IV along with CLEAN to produce the first total intensity images of the 230 GHz core in M87. RML algorithms find an image that maximizes an objective function composed of a likelihood term comparing the image to data, and regularizer terms that penalize or favor certain image features. In this work, we use the RML method implemented in the `eht-imaging`[6] software library (Chael

[3] `http://www.aips.nrao.edu`

[4] `https://casa.nrao.edu`

[5] `GPCAL` is a new automated pipeline written in Python and based on AIPS and the CLEAN imaging software `Difmap`. `GPCAL` adopts a similar calibration scheme to `LPCAL` but allows users to (i) fit the D-term model to multiple calibrators simultaneously and (ii) use more accurate linear polarization models of the calibrators for D-term estimation. In this paper, it is mainly used to complement the `LPCAL` analysis of the M87 data (Appendix G.3 and K) and the D-term estimation using calibrators (Appendix J).

[6] `https://github.com/achael/eht-imaging`

et al. 2016, 2018) to solve for images in both total intensity and linear polarization from the EHT data. Like the CLEAN-based methods, `eht-imaging` does not solve for Stokes $\mathcal{V}$. Details on the specific imaging methods in `eht-imaging` used in the reconstructions presented in this work can be found in Appendix C.2.

Imaging as posterior exploration is carried out using two independent Markov chain Monte Carlo (MCMC) schemes: DMC and Themis. Both codes simultaneously explore the posterior space of the full Stokes image (including Stokes $\mathcal{V}$) alongside the complex gains and leakages at every station; station gains are permitted to vary independently on every scan, while leakage parameters are modeled as constant in time throughout an observation. We provide more detailed model specifications for both codes in Appendix C.3 and in separate publications (D. W. Pesce et al. *in prep.*, A. E. Broderick et al. *in prep.*).

Hereafter we often refer to `eht-imaging`, `polsolve`, and `LPCAL` methods as *imaging* methods/pipelines and to DMC and Themis methods as *posterior exploration* methods/pipelines.

4.2. *Leakage and gain calibration strategy*

In the imaging methods we divide the polarimetric calibration procedure for EHT data into two steps. We first calibrate the stations with an intra-site partner (ALMA–APEX, SMA–JCMT) using the assumption that sources are unresolved on intra-site baselines, where the brightness distribution can be approximated with a simple point source model. In the imaging pipelines we apply the D-terms for ALMA, APEX and SMA to the data prior to polarimetric imaging and D-term calibration of the remaining stations. Baselines to the JCMT (redundant with SMA baselines) are subsequently removed from the data sets, to reduce complications from single-polarization station handling. In the second step we perform simultaneous imaging of the source brightness distribution and D-term calibration of stations for which only long source-resolving baselines are available. In contrast, the posterior exploration pipelines do not use the D-terms derived using the zero-baseline approach and instead solve for all D-terms (and station gains) starting with the base data product described in Section 3.2.

The point source assumption adopted in the imaging method zero-baseline D-term calibration step is an extension to the intra-site redundancies already exploited in the EHT network calibration (Paper III), allowing us to obtain a model-independent gain calibration for ALMA, APEX, SMA, and JCMT. For an unresolved, slowly evolving source we can assume the true parameters of the coherency matrix ρ_{jk} in Equation 6, to be constant throughout a day of observations, since very low spatial frequencies $\mathbf{u}$ are sampled, $\rho_{jk} \approx \rho_{jk}(\mathbf{u} = 0)$. Hence, only four intrinsic visibility components of ρ_{jk} per source and four complex D-terms (two for each station) need to be determined from all the data on an available baseline.

We fit the D-terms of ALMA, APEX, JCMT and SMA for each day using the multi-source feature of `polsolve`, combining band-averaged observations of multiple sources (3C 279, M87, J1924–2914, NRAO 530, 3C 273, 1055+018, OJ287 and Cen A as shown in Appendix D) on each day in one single fit per band. The results of these fits per station, polarization, day, and band are presented in Figure 2 (left panel), where average D-terms across all days are plotted for each station, band and polarization, with the standard deviation estimated from the dispersion of the D-terms across all days. In Appendix D, we provide tables with D-term values and further discuss the time and frequency dependence of D-terms and JCMT single polarization handling. In Appendix D we also present several validation tests of our zero-baseline D-term estimation method carried out to motivate the use of band-averaged data products, comparisons to independent polarimetric source properties measured from simultaneous interferometric-ALMA (Goddi et al in prep 2020) and near-in-time interferometric-SMA leakage estimates, and comparisons to results from a model fitting approach.

In addition to zero-baseline D-term calibration in the imaging pipelines, we also account for residual station-based amplitude gain errors by calibrating the data to pre-determined best Stokes $\mathcal{I}$ images of chosen sources. Given the extreme resolving power of the EHT array, all available calibrators are resolved on long baselines. We must therefore select sources best imaged with our EHT array configuration with high fidelity: compact non-variable sources with sufficient (u, v) coverage for imaging. This leaves us four targets in our EHT 2017 observations fitting these criteria: M87, 3C 279, J1924–2914 and NRAO 530. M87 and 3C 279 Stokes $\mathcal{I}$ have been published (Paper I – Paper VI, Kim et al. 2020). Final Stokes $\mathcal{I}$ images for the Sgr A* calibrators NRAO 530 and J1924–2914 will be presented in upcoming publications (S. Issaoun et al. *in prep.*, S. Jorstad et al. *in prep.*) but the best available preliminary images are used to self-calibrate our visibility data for D-term comparisons in this paper.

Although multiple imaging packages and pipelines were utilized in the Stokes $\mathcal{I}$ imaging process, the resulting final 'fiducial' images from each method are highly consistent at the EHT instrumental resolution (e.g., Paper IV, figure 15). We therefore selected a set of Stokes $\mathcal{I}$ images for self-calibration from the RML-based `SMILI` imaging software pipeline (Akiyama et al. 2017a,b, Paper IV). The images we use for self-calibration are at `SMILI`'s native imaging resolution ($\sim$ 10 μas), which provide the best fits to the data, and are not convolved with any restoring beam. We self-calibrate our visibility data to these images, therefore accounting for station gain residual variations in the data that make imaging challenging. Via self-calibration, the data sets used in

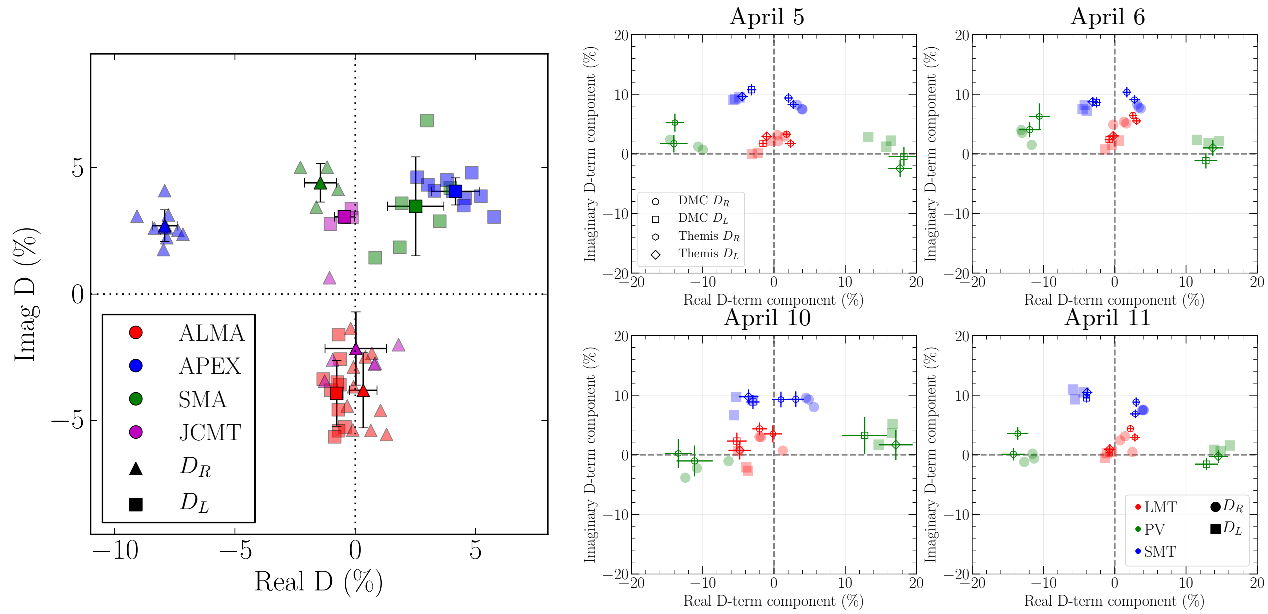

Figure 2. *Left panel:* D-term estimates for ALMA, APEX, JCMT, and SMA from `polsolve` multi-source intra-site baseline fitting; one point per day and band (low and high) for each station across the EHT 2017 campaign. Both polarizations are shown for ALMA and APEX per day, one for JCMT and SMA per day due to JCMT polarization setup limitations. Station averages across days, high bands and low bands are shown with error bars. The depicted D-terms are provided in tabulated form in Appendix D. *Right panels*: Fiducial D-terms for LMT, PV, and SMT derived via leakage calibration through polarimetric imaging methods and posterior modeling of M87 observations. We depict fiducial D-terms per day, where each point corresponds to one station, polarization, and method. Filled symbols depict D-terms from imaging methods and symbols for posterior exploration methods have errorbars corresponding to the 1-σ standard deviations estimated from posterior distributions of the resulting D-terms.

polarimetric imaging methods become easier to image and allow methods to focus on accurate reconstructions of polarimetric Stokes $\mathcal{Q}$ and $\mathcal{U}$ brightness distributions and D-term estimation.

Preliminary D-terms estimated via imaging methods are reported in Appendix F. To quantify the agreement (or distance in the complex plane) between D-term estimates from different methods we calculate L_1 norms. L_1 norms averaged over left and right (also real and imaginary) D-term components, over all stations and over time are less than 1% for each pair of imaging methods (see Figure 20 in Appendix E). The mean values of D-term posteriors from the posterior exploration methods correlate well with D-terms estimates by imaging methods. For each combination of imaging and posterior exploration method the station averaged L_1 norms range from 1.5% to 1.89%. Any residual leakage for stations with a co-located partner estimated by `LPCAL`[7], and both

[7] The residual leakage estimated by `LPCAL` is due to `LPCAL` being unable to fix D-terms of specific stations to be certain values. Thus, `LPCAL` obtains solutions for those stations and the non-zero D-terms indicate that there may either be possible residual leakage after zero-baseline fitting or uncertainties in `LPCAL` estimates originating from e.g., a breakdown of the similarity approximation.

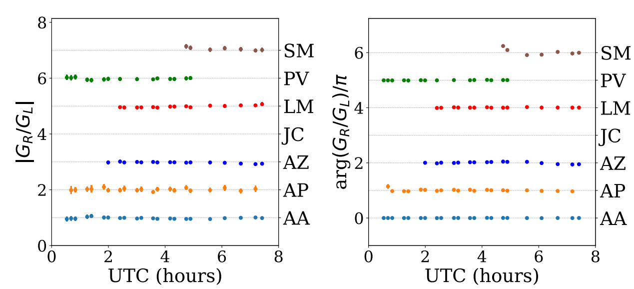

Figure 3. Amplitudes (left) and phases (right) of the ratio of R to L station gains from the DMC fit to M87 April 11 low-band data. Individual station gain ratios are offset vertically for clarity, with the dashed horizontal lines indicating a unit ratio for each station. Note that JCMT only observes one polarization at a time, and so provides no constraints on gain ratios. We see that the assumption made by the three imaging pipelines and one posterior exploration pipeline (Themis) – namely, that the right- and left-hand gains are equal for all stations at all times – largely seems to hold. The behavior in this plot is representative of that seen across days and bands.

posterior exploration methods is small. The gain calibration (see Equation 4) for DMC is shown in Figure 3. As expected, the assumption made by all of the imaging pipelines and one of the posterior exploration pipelines (Themis), that right- and left-hand gains are equal for all stations at all times, holds. For verification purpose, we also estimate D-terms using data of calibrating sources. We find that the D-terms derived by polarimetric imaging of other sources are consistent with those of M87. The results are presented in Appendix J. Finally, our estimated SMT D-terms are similar to those computed previously using EHT observations of Sgr A* (Johnson et al. 2015).

4.3. *Parameter Surveys and Validation on Synthetic Data*

Each imaging and leakage calibration method has free parameters that must be set by the user before the optimization or posterior exploration takes place. Some of these parameters (e.g., field of view, number of pixels) are common to all imaging methods, but many are unique to each method (e.g., the sub-component definitions in `LPCAL` or `polsolve`, or the regularizer weights in `eht-imaging`). Often in VLBI imaging, these free parameters are simply set by the user given their experience on similar data sets, or based on what appears to give a good fit to the data free of noticeable imaging artifacts. In this work, we follow Paper IV in choosing the method parameters we use in our final image reconstructions more objectively by surveying a portion of the parameter space available to each method.

We perform surveys over the different free parameters available to each imaging method and attempt to chose an optimal set of parameters based on their performance faithfully recovering the source structure and input D-terms from our synthetic data models. The parameter set that performs best on the synthetic data for each method is considered our "fiducial" parameter set for imaging M87 with that method. The corresponding images reconstructed from various data sets using these parameters are the method's "fiducial images".

The synthetic data sets we used for scoring the imaging parameter combinations consist of six synthetic EHT observations using the M87 April 11 equivalent low band (u, v) coverage. The source structure models used in the six sets vary from complex models for source structure generated using general relativistic magnetohydrodynamics (GRMHD) simulations of M87's core and jet base (Models 1 and 2 from Chael et al. 2019) to simple geometrical models (a filled disk, Model 3, and simple rings with differing EVPA patterns, Models 4-6). The synthetic source models have varying degrees of fractional polarization and diverse EVPA structures. The mock data source models blurred to the EHT nominal resolution are displayed in the first column of Figure 4.

All M87 synthetic data sets were generated using the synthetic data generation routines in `eht-imaging`. We followed the synthetic data generation procedure in Appendix C.2 of Paper IV, but with models featuring complex polarization structure. The synthetic visibilities sampled on EHT baselines are corrupted with thermal noise, phase and gain offsets, and polarimetric leakage terms. Mock D-terms for the SMT, LMT, and PV stations were chosen to be similar to those found by the initial exploration of the M87 EHT 2017 data reported in Appendix F. Random residual D-terms for ALMA, APEX, JCMT, and SMT (reflecting possible errors in the zero-baseline calibration procedure) were drawn from normal distributions with 1% standard deviation. After generation, the phase and amplitude gains in the synthetic data were calibrated for use in imaging pipelines in the same way as the real M87 data; that is, they were self-calibrated to a Stokes $\mathcal{I}$ image reconstructed via the `SMILI` fiducial script for M87 developed in Paper IV.

In Figure 4, we present our fiducial set of images (in a uniform scale) from synthetic data surveys carried within each method. In each panel we report a correlation coefficient $\langle I \cdot I_0 \rangle$ between recovered Stokes $\mathcal{I}$ and the ground truth $\mathcal{I}_0$ images,

$$\langle I \cdot I_0 \rangle = \frac{\langle\, (\mathcal{I} - \overline{\mathcal{I}})(\mathcal{I}_0 - \overline{\mathcal{I}}_0)\, \rangle}{\sqrt{\langle\, (\mathcal{I} - \overline{\mathcal{I}})^2\, \rangle}\sqrt{\langle\, (\mathcal{I}_0 - \overline{\mathcal{I}}_0)^2\, \rangle}}. \qquad (8)$$

This reflects the dot product of the two mean-subtracted images when treated as unit vectors. We also calculate a correlation coefficient for the reconstructed linear polarization image $p \equiv \mathcal{Q} + i\mathcal{U}$,

$$\langle \vec{P} \cdot \vec{P}_0 \rangle = \frac{\mathrm{Re}[\langle\, p\, p_0^* \,\rangle]}{\sqrt{\langle\, p\, p^* \,\rangle}\sqrt{\langle\, p_0\, p_0^* \,\rangle}}. \qquad (9)$$

The real part is chosen to measure the degree of alignment of the polarization vectors $(\mathcal{Q}, \mathcal{U})$. In both cases, images are first shifted to give the maximum correlation coefficient for Stokes $\mathcal{I}$. Because Stokes $\mathcal{I}$ image reconstructions are tightly constrained by an a priori known total image flux, the Stokes $\mathcal{I}$ correlation coefficients are mean subtracted to increase the dynamic range of the comparison. This introduces a field of view dependence to the metric, as only spatial frequencies above (field of view)$^{-1}$ are considered; up to the beam resolution. There is no such dependence in the linear polarization coefficient which is not mean subtracted.

The correlation is equally strong independently of the employed method. The polarization structure is more difficult to recover for models with high or complex extended polarization (Models 1 and 2) for which correlation of the recovered polarization vectors is strong to moderate. This seems to be independent of the method as well. In Figure 5 we present a uniform comparison of the recovered D-terms and the ground truth D-terms for all synthetic data sets and methods. For all methods the recovered D-terms show a strong correlation with the model D-terms. To quantify the agreement (or distance in the complex plane) between D-term estimates

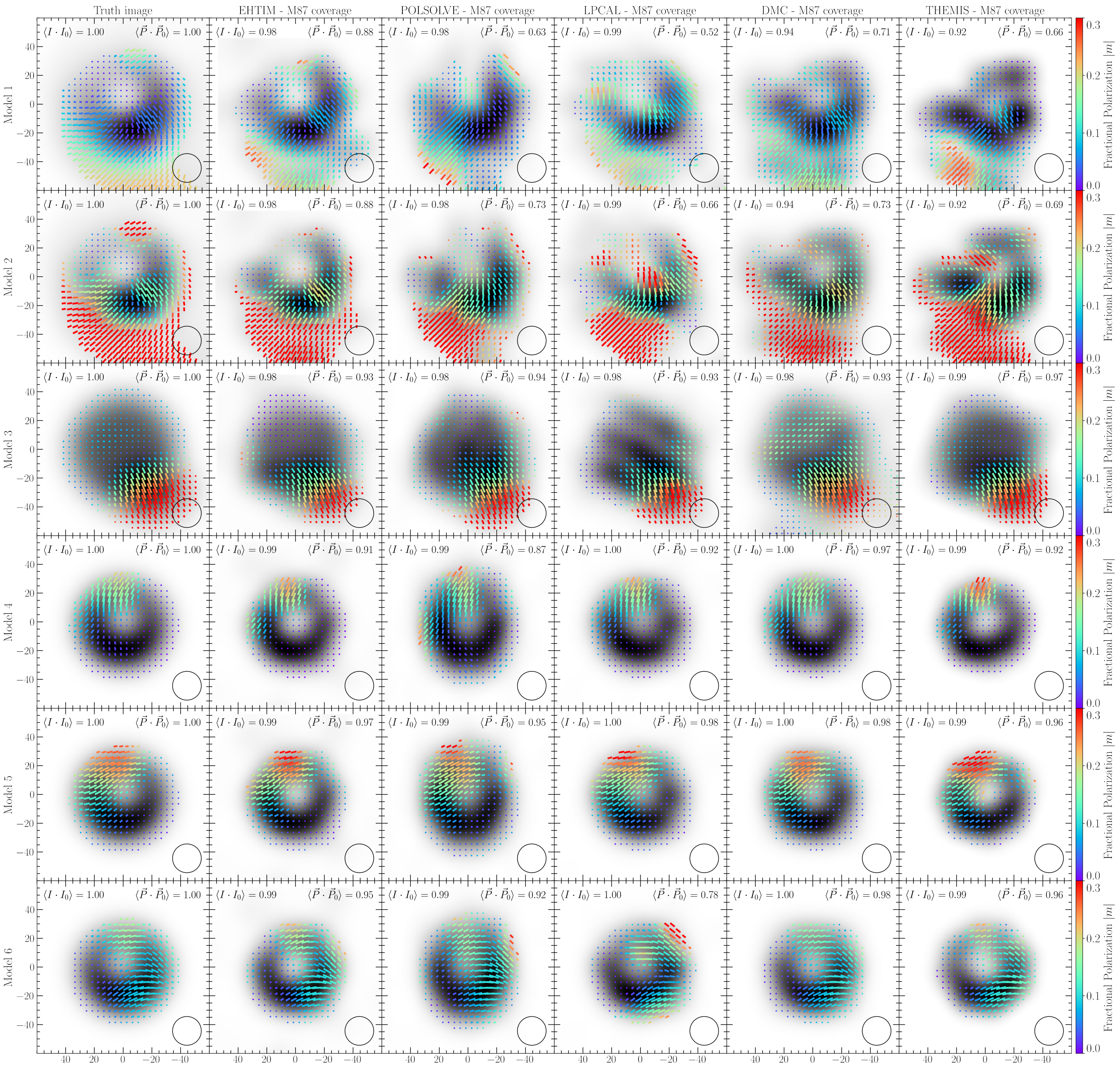

Figure 4. Fiducial images from synthetic data model reconstructions using M87 April 11 low band (u, v) coverage. Rows from top to bottom correspond to six different synthetic data sets. Columns from left to right show ground truth synthetic image (column 1) and the best image reconstructions by each method (columns 2-6). The polarization tick length reflects total linear polarization, while the color reflects fractional polarization from 0 to 0.3. The normalized overlap is calculated against the respective ground truth image, and in the case of the total intensity it is mean-subtracted.

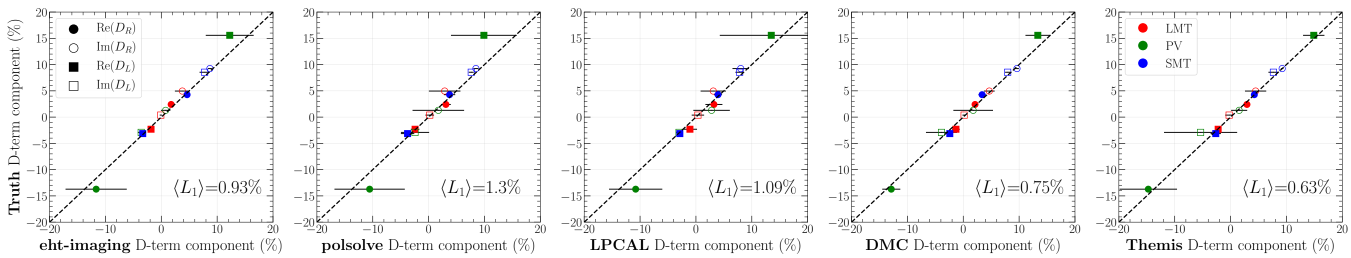

Figure 5. A comparison of LMT, SMT, and PV D-term estimates to ground truth values in the synthetic data sets 1 through 6 (shown in Figure 4). Each panel shows correlation of the estimated and the truth D-terms for a single method. Each data point in each panel depicts an average and standard deviation for each D-term estimate derived from the six synthetic data sets. The norm $L_1 \equiv |D - D_{\rm Truth}|$ is averaged over left, right, real, and imaginary components of the D-terms and over all shown EHT stations. Notice that each method recovers the ground truth D-terms to within $\sim 1\%$, on average.

and the ground truth values $D_{\rm Truth}$ in each approach, we calculate the $L_1 \equiv |D_i - D_{\rm Truth}|$ norm, where D_i is a D-term component derived within a method i. Overall, for the fiducial set of parameters the agreement between the ground truth and the recovered D-terms in synthetic data measured using the L_1 norm is $\leq 1.3\%$ on average (when averaging is done over stations, D-term components, and models). The reported averaged L_1 norms give us a sense of the expected discrepancies in D-terms between employed methods for their fiducial set of parameters. However, we notice again that the discrepancies do depend on source structure. For example, in models with no polarization substructure (e.g., Model 3) all methods had difficulty in recovering D-terms for PV (visible as large error bars for the antenna), a station forming only very long baselines on a short (u, v) track. If we exclude PV from the L_1 metrics the expected L_1 norms for LMT and SMT alone for all methods are $L_1 \sim 0.6 - 0.8\%$ when averaged over models.

5. RESULTS

5.1. *Fiducial Polarimetric Images of M87*

In Figure 6, we present the fiducial M87 linear polarimetric images produced by each method from the low band data on all four observing days. Unless otherwise explicitly indicated, we focus on displaying low-band results in the main text, as the high-band results, very close in observing frequency, are consistent and do not significantly deviate from low-band results (see Appendix I for a detailed comparison of low and high band results). The different imaging reconstruction methods have different intrinsic resolution scales; for instance, the CLEAN reconstruction methods model the data as an array of point sources, while the RML and MCMC methods have a resolution scale set by the pixel size. In Figure 6, we display the fiducial images from each method at the same resolution scale by convolving each with a circular Gaussian kernel with a different FWHM. The FWHM for each method is set by maximizing the normalized cross-correlation of the blurred Stokes $\mathcal{I}$ image with the April 11 "consensus" image presented in Figure 15 of Paper IV. The blurring kernel FWHMs selected by this method are 19μas for `eht-imaging`, DMC, and Themis, 20 μas for `LPCAL`, and 23 μas for `polsolve`.

The fiducial images from each method are broadly consistent with those from the preliminary imaging stage shown in Appendix F in Figure 21. The M87 emission ring is polarized only in its south–west region and the fractional polarization at $\approx 20\,\mu$as resolution is at the level of about 15%. The residual rms in linear polarization (as estimated from the CLEAN images) is between 1.10–1.30 mJy/beam in all epochs, which implies a polarization dynamic range of $\sim$10. The nearly azimuthal EVPA pattern is a robust feature evident in all our reconstructions. The images show slight differences in the polarization structure between the first two days, April 5/6 and the last two, April 10/11. Notably, the south part of the ring appears less polarized on the later days. This evolution in the polarized flux density is consistent with the evolution in the Stokes $\mathcal{I}$ image apparent in the underlying closure phase data (Paper III, Figure 14; Paper IV, Figure 23). However, as with the Stokes $\mathcal{I}$ image, the structural changes in the polarization images with time over this short timescale (6 days $\approx$ 16 GM/c^3) are relatively small and it is difficult to disentangle which differences in the polarized images are robust and which are influenced by differences in the interferometric (u, v) coverage between April 5 and April 11. (Paper IV, Section 8.3)

In Figure 7, we show the simple average of the five equivalently blurred fiducial images (one per method) for each of the four observed days. The averaging is done independently for each Stokes brightness distribution. These averaged images are consistent with the EHT closure traces as shown in Figure 13 in Appendix B. We adopt the images in Figure 7 as a conservative representation of our final M87 polarimetric imaging results.

5.2. *Azimuthal distribution of the polarization brightness*

While the overall pattern of the linearly polarized emission from M87 is consistent from method to method, the details of the emission pattern can depend sensitively

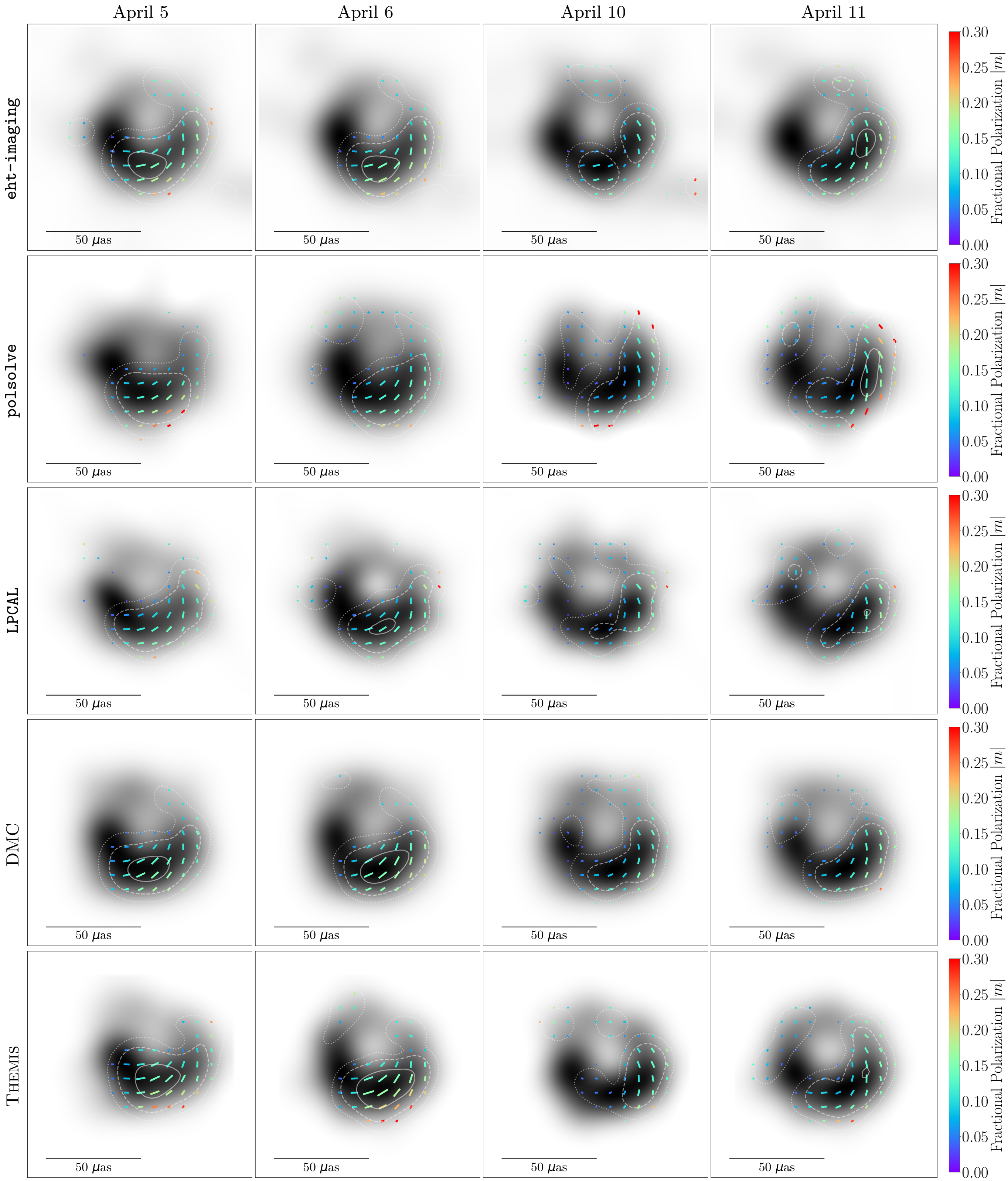

Figure 6. Fiducial M87 images produced by five independent methods. Results from all imaging and posterior exploration pipelines are shown on all four M87 low-band observation days (low and high band results are consistent, see Appendix I). Total intensity is shown in grayscale, polarization ticks indicate the EVPA, the tick length indicates linear polarization intensity magnitude, and color indicates fractional linear polarization. The tick length is scaled according to the polarized brightness without renormalization to the maximum for each image. The contours mark linear polarized intensity. The solid, dashed, and dotted contour levels correspond to linearly polarized intensity of 20, 10, and 5 μJy/μas^2. Cuts were made to omit all regions in the images where Stokes $\mathcal{I} < 10\%$ of the peak flux density and $p < 20\%$ of the peak polarized flux density. The images are all displayed with a field of view of 120 μas, and all images were brought to the same nominal resolution by convolution with the circular Gaussian kernel that maximized the cross-correlation of the blurred Stokes $\mathcal{I}$ image with the consensus Stokes $\mathcal{I}$ image of Paper IV.

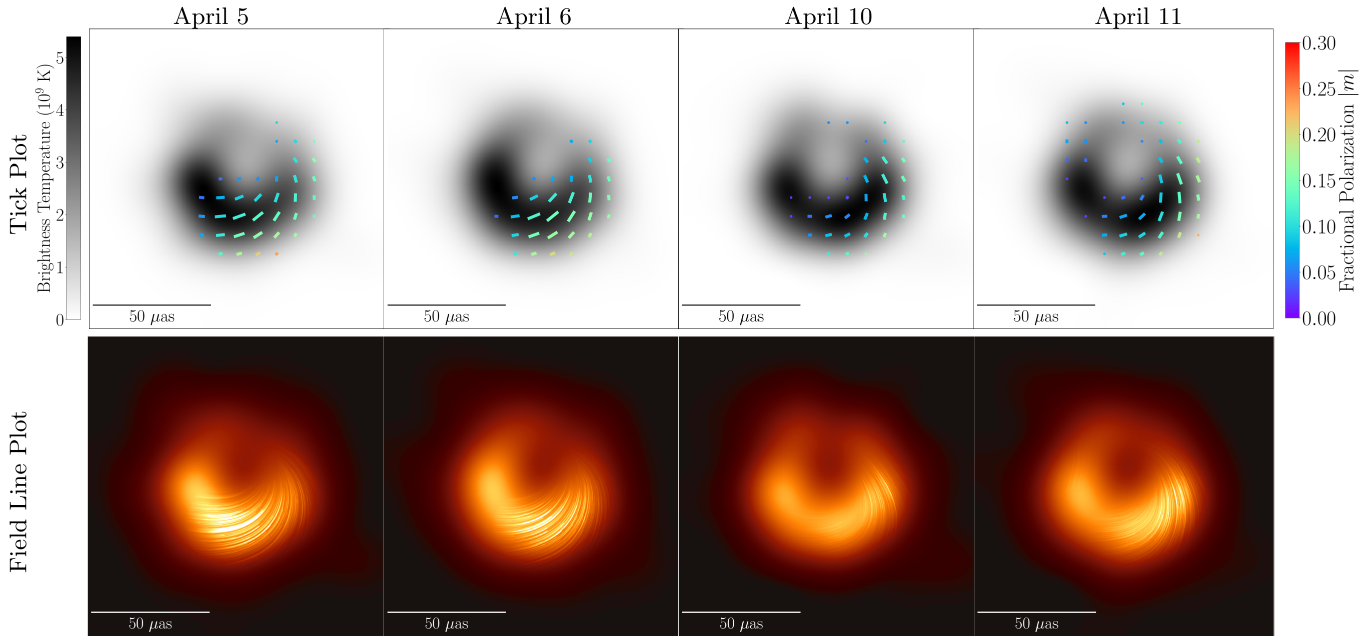

Figure 7. Fiducial M87 average images per day produced by averaging results from our five methods (see Figure 6). Method-average images for all four low-band M87 observation days are shown, from left to right (low and high band results are consistent, see Appendix I). We employ here two visualization schemes (top and bottom) to display our four method-average images. The images are all displayed with a field of view of 120 μas, and all images were brought to the same nominal resolution by convolution with the circular Gaussian kernel that maximized the cross-correlation of the blurred Stokes $\mathcal{I}$ image with the consensus Stokes $\mathcal{I}$ image of Paper IV. *Top*: Total intensity, polarization fraction, and EVPA are plotted in the same manner as in Figure 6. *Bottom*: Polarization "field lines" plotted atop an underlying total intensity image. Treating the linear polarization as a vector field, the sweeping lines in the images represent streamlines of this field and thus trace the EVPA patterns in the image. To emphasize the regions with stronger polarization detections, we have scaled the length and opacity of these streamlines as the square of the polarized intensity. This visualization is inspired in part by Line Integral Convolution (LIC; Cabral & Leedom 1993) representations of vector fields, and it aims to highlight the newly added polarization information on top of the standard visualization for our previously published Stokes $\mathcal{I}$ results (Paper I; Paper IV).

on the remaining statistical uncertainties in our leakage calibration. In addition, the different assumptions and parameters used in each imaging method affect the recovered polarized intensity pattern, introducing an additional source of systematic uncertainty in our recovered images. In this section, we assess the consistency of the recovered polarized images across different D-term calibration solutions within and between methods.

We explore the consistency of our image reconstructions against the uncertainties in the calibrated D-terms by generating a sample of 1000 images for each method, each generated with a different D-term solution. For the imaging methods, we define complex normal distributions for each D-term based on the results of Figure 2 and reconstruct images after calibrating to each set of random D-terms. This procedure is explained in detail in Appendix H. For the posterior exploration methods we simply draw 1000 images from the posterior for each observing day.

For each method's set of 1000 image samples covering a range of D-term calibration, we study the azimuthal distribution of the polarization brightness (p) and EVPA (χ) by performing intensity-weighted averages of these quantities over different angular sections along the ring. The width of the angular sections used in the averaging is set to $\Delta\varphi = 10°$ and the averages are computed from a position angle $\varphi = 0°$ to $\varphi = 360°$, in steps of 1°.

Comparing angular averages of these quantities with a small moving window $\Delta\varphi$ avoids spurious features from e.g., the different pixel scales used in the different image reconstruction methods. The pixel coordinates of the image center are estimated (for each method) from the peak of the cross-correlation between the $\mathcal{I}$ images and the representative images of M87 used in the self-calibration. To avoid the effects of phase wrapping in the averaging (which biases the results for values of χ around $\pm$ 90°), the quantity $\langle\chi\rangle$ is computed coherently within each angular section, i.e., the averages are defined as

$$\langle p\rangle \equiv \frac{\langle \mathcal{I}\sqrt{\mathcal{Q}^2+\mathcal{U}^2}\rangle}{\langle\mathcal{I}\rangle}, \tag{10}$$

$$\langle\chi\rangle \equiv \frac{1}{2}\arctan\left(\frac{\langle\mathcal{Q}\times\mathcal{I}\rangle}{\langle\mathcal{U}\times\mathcal{I}\rangle}\right). \tag{11}$$

In Figure 8, we show the histograms of these quantities for two days, April 5 and April 11, as a function of the orientation of the angular section used in the averaging (i.e., the position angle). We consider these two days because they have the best (u, v) coverage and span the full observation window; these results will thus include any effects of intrinsic source evolution in the recovered parameters. From Figure 8, it is evident that the difference in $\langle p \rangle$ between methods is larger than the widths of the $\langle p \rangle$ histograms in each method. This means that the effects related to the residual instrumental polarization, giving rise to the dispersion seen in the histograms, are smaller than the artifacts related to the deconvolution algorithms. In other words, the $\langle p \rangle$ images are limited by the image fidelity due to the sparse (u, v) coverage rather than by the D-terms.

Even though there are differences among methods in the p azimuthal distribution, some features are common to all our image reconstructions. On one hand, the peak in the polarization brightness is located near the south-west on April 5 (at a position angle of $199 \pm 11°$, averaged among all methods) and close to the west on April 11 (position angle of $244 \pm 10°$). That is, the polarization peak appears to rotate counter-clockwise between the two observing days (see dotted lines in Fig. 8). In any case, the region of high polarization brightness is relatively wide, covering a large fraction of the southern portion of the image (position angles from around 100° to 300°).

Regarding the azimuthal distribution of $\langle \chi \rangle$, all methods produce very similar values in the part of the image with highest polarized brightness (the south-west region, between position angles of 180° and 270°) . The EVPA varies almost linearly, from around $\langle \chi \rangle = -80°$ (close to the South) up to around $\langle \chi \rangle = 30°$, close to the east. The EVPAs on April 11 are slightly higher (i.e., rotated counter-clockwise) compared to those on April 5. This difference is clearly seen for `eht-imaging`, `polsolve`, and THEMIS, though the difference is smaller for DMC and `LPCAL`. We notice, though, that the differences in the EVPAs between days could also be affected by small shifts in the estimates of the image center at each day. Outside of the region with high polarization, the EVPA distributions for all methods start to depart from each other. There is a hint of $\langle \chi \rangle \sim 0°$ in the northern region (i.e., position angles around $0 - 50°$), which is clearly seen in `polsolve` and `LPCAL` in both days (though less clearly in the other methods).

The discrepancies in EVPA among all methods only appear in the regions with low brightness (i.e., around the northern part of the ring). Therefore, polarization quantities defined from intensity-weighted image averages, discussed in the next sections, will be dominated by the regions with higher brightness, for which all methods produce similar results. Image-averaged quantities are somewhat more robust to differences in the calibration and imaging algorithms, though they are not immune to systematic errors.

5.3. *Image-averaged quantities*

In comparing polarimetric images of M87, we are most interested in identifying acceptable ranges of three image-averaged parameters that are used to distinguish between different accretion models in Paper VIII: the net linear polarization fraction of the image $|m|_{\rm net}$, the average polarization on the image at $20\,\mu$as resolution $\langle m \rangle$, and the $m = 2$ coefficient of the azimuthal mode decomposition of the polarized brightness β_2. These parameters are defined below.

First, the net linear polarization fraction of the image is

$$|m|_{\rm net} = \frac{\sqrt{\left(\sum_i Q_i\right)^2 + \left(\sum_i U_i\right)^2}}{\sum_i I_i}. \quad (12)$$

ALMA measured $|m|_{\rm net} = 2.7\%$ on April 11(Goddi et al in prep 2020), but this measurement includes emission at large scales outside of the $120\,\mu$as field of view of the EHT images. We also consider the intensity-weighted average polarization fraction across the resolved EHT image:

$$\langle |m| \rangle = \frac{\sum_i \sqrt{Q_i^2 + U_i^2}}{\sum_i I_i}. \quad (13)$$

The value of $\langle |m| \rangle$ is determined by the intensity of the polarized emission at each point in the image, and it is thus sensitive to the resolution of the image and the choice of restoring beam. Specifically, images restored with beams of larger FWHM will tend to be more locally depolarized and thus have lower $\langle |m| \rangle$ than images restored with beams of smaller FHWM. In contrast, the integrated polarization fraction $|m|_{\rm net}$ is insensitive to convolution.

We quantify the polarization *structure* with a decomposition into azimuthal modes. In particular Paper VIII considers the complex amplitude β_2 of the $m = 2$ mode defined in Palumbo et al. (2020), who found this mode to be the most important in distinguishing different modes of accretion from 230 GHz images produced by different GRMHD simulations. The β_2 azimuthal mode decomposition coefficient is defined as

$$\beta_2 = \frac{1}{I_{\rm ring}} \int_{\rho_{\rm min}}^{\rho_{\rm max}} \int_0^{2\pi} P(\rho, \varphi)\, e^{-2i\varphi}\, \rho\, d\varphi\, d\rho, \quad (14)$$

where (ρ, φ) are polar coordinates in the image plane, and $I_{\rm ring}$ is the Stokes $\mathcal{I}$ flux density along the ring between the minimum radius $\rho_{\rm min}$ and the maximum radius $\rho_{\rm max}$. Since our imaging methods recover no significant extended flux density off the main ring, we take $\rho_{\rm min} = 0$ and extend $\rho_{\rm max}$ to encompass the full image field of view, and $I_{\rm ring}$ is equal to the total Stokes $\mathcal{I}$ intensity in the image.

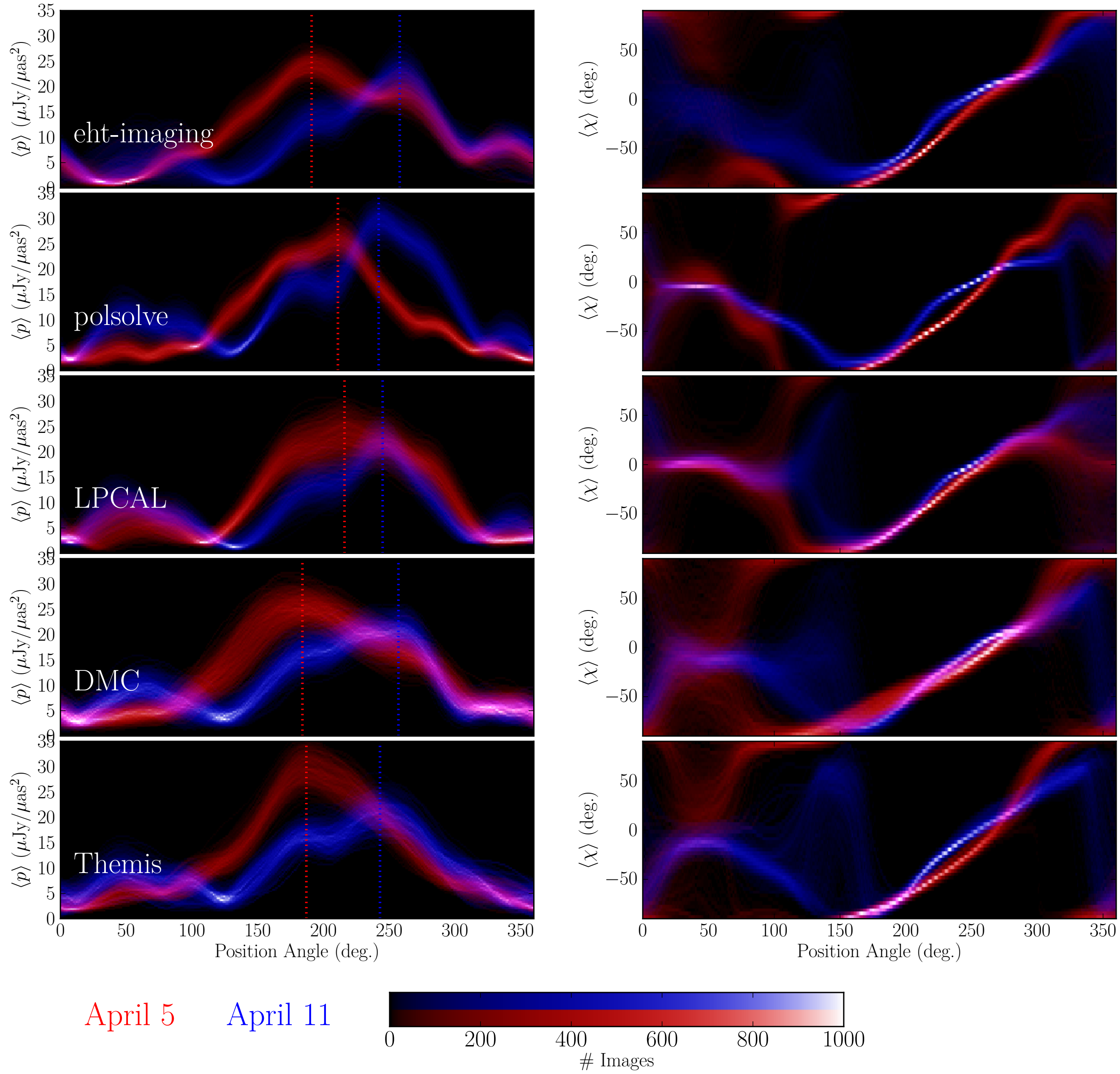

Figure 8. Histograms of the azimuthal distributions of polarized intensity (left) and EVPA (right) obtained from the Monte Carlo D-term analysis with all imaging and posterior exploration methods. These quantities are estimated as the intensity-weighted averages within an angular section of a width of 10 deg. The position angle is measured counter-clockwise, starting from North. The position angles with highest average polarization brightness are marked with dotted lines for each method and day.

Note that both the amplitude $|\beta_2|$ and phase $\angle\beta_2$ depend on the choice of image center, image resolution, and restoring beam size. In the comparisons that follow, we convolve images from each method with Gaussian beams of the fiducial FWHMs specified in Section 5.1. Furthermore, we center each reconstruction by finding the pixel offset that maximizes the cross-correlation between the blurred Stokes $\mathcal{I}$ image and the April 11 consensus Stokes $\mathcal{I}$ image from Paper IV. In general, we find these offsets to be small and our results do not change significantly if we do not apply any centering procedure in calculating β_2 from our reconstructed images.

From the sets of 1000 images generated from each method to explore variations of the image structure with the D-term solution, we compute distributions of each of these key metrics used in Paper VIII for theoretical interpretation – $|m|_{\rm net}, \langle m\rangle, |\beta_2|$, and $\angle\beta_2$.

These distributions are summarized in Figure 9, which displays the mean points and 1-σ error bars for all four methods on both April 5 and 11. We present a more complete look at these distributions with histograms for each quantity from each method/day in Appendix H, Figures 25 and 26.

On the same observation day, the distributions of $|m|_{\rm net}$ appear consistent between most pairs of reconstruction methods, with some notable exceptions. Many of the distributions of $|m|_{\rm net}$ peak around the ALMA measured value of 2.7%, but the `LPCAL` distributions on both days and the `eht-imaging` distributions on April 11 are peaked closer to 1%. The distributions of $\langle|m|\rangle$ are peaked between 6 and 11 % for all five methods across both days. On both days, the $\langle|m|\rangle$ distributions for `eht-imaging`, DMC, and Themis are peaked at values 2–3% higher than the `LPCAL` or `polsolve` distributions. This systematic shift may indicate residual issues with bringing the imaging methods to the same resolution scale; in particular, the same circular Gaussian kernel was used to blur Stokes $\mathcal{I}, \mathcal{Q}$ and $\mathcal{U}$ in each method, while the intrinsic resolution of the reconstruction in $\mathcal{Q}$ and $\mathcal{U}$ may be lower than in total intensity. In each method, there appears to be a decrease in $\langle|m|\rangle$ of $\approx 1-2\%$ on between April 5 and 11.

The mean of the amplitude $|\beta_2|$ distribution is peaked between 0.04 and 0.07 for all methods on both days; however the results from `eht-imaging`, DMC, and Themis appear larger in amplitude on both days than the corresponding distributions for `polsolve` and `LPCAL`. Again, since $|\beta_2|$ is sensitive to the restoring beam size, this may be due to residual errors in bringing the polarized images to the same resolution scale. Similarly to the distributions of $\langle|m|\rangle$, there are indications of a shift downward in $|\beta_2|$ by an absolute value of ≈ 0.01 in all four methods between April 5 and 11. The distributions of the phase $\angle\beta_2$ are consistent between most pairs of methods with no obvious systematic difference between the sub-component methods (`LPCAL`, `polsolve`) and those that use a continuous image representation (`eht-imaging`, DMC, and Themis). Furthermore, there is no apparent systematic difference in the $\angle\beta_2$ results between April 5 and 11.

To score different accretion models from GRMHD simulations against constraints from the EHT data, Paper VIII uses a range for each quantity that incorporates both the uncertainties in the parameters from the D-term calibration process (the error bars for each method in Figure 9) and the systematic uncertainty across imaging methods (the scatter in the points). The final ranges used in Paper VIII for each parameter were set by taking the minimum/maximum of the ten mean values minus/plus 1-σ across both days and all methods. These parameter ranges are denoted by colored bands in Figure 9 and presented in Table 2.

Table 2. Final parameter ranges for the quantities used in scoring accretion/jet models in Paper VIII.

Parameter	Min	Max
$\lvert m\rvert_{\rm net}$	1.0%	3.7 %
$\langle\lvert m\rvert\rangle$	5.7 %	10.7 %
$\lvert\beta_2\rvert$	0.04	0.07
$\angle\beta_2$	-163 deg	-127 deg

Note – The ranges are taken from the bands plotted in Figure 9 incorporating the $\pm1\sigma$ error from each method's D-term calibration survey.

6. DISCUSSION

We discuss several important effects in the polarimetric emission from M87 that are relevant for our analysis of the 230 GHz linear polarization structure in this work.

Figure 8 demonstrates variability in the total intensity and polarimetric images of M87 between April 5 and April 11. It is unlikely that the polarimetric variability in the reconstructed images is due to the different (u, v) coverages on different days. The changes in the polarimetric images are consistent with signatures of the source intrinsic variability noticed in the VLBI data themselves. Variability of specific, calibration insensitive, EHT VLBI data products are introduced and discussed in Appendix B.

The total flux from M87 inner arc-second measured on intra-site baselines (ALMA–APEX) and measured by ALMA-only is $F \sim 1.2$ Jy; this is a factor of 2 higher than total flux measured in the ring visible on EHT scales. Given that the net fractional polarization measured in the EHT images ($|m| \sim 1-3.7\%$) is consistent with that measured on larger (arc-second) scales $|m| \sim 2.7\%$ (Goddi et al in prep 2020), the net fractional polarization of any other emission component(s) in ALMA field-of-view should be comparable to that of the M87 ring.

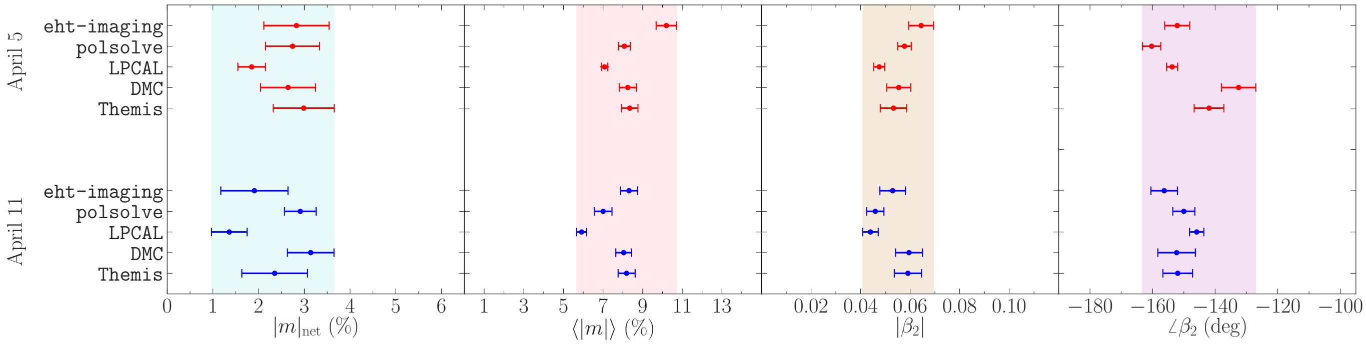

Figure 9. Summary of the results for the key quantities used in Paper VIII for each method on both April 5 and 11. From left to right, the quantities are the integrated net polarization $|m|_{\rm net}$ (Equation 12), the average polarization fraction $\langle|m|\rangle$, and the amplitude $|\beta_2|$ and phase $\angle\beta_2$ of the $m=2$ azimuthal mode of the complex polarization brightness distribution (Equation 14). The shaded bands show the consensus ranges (Table 2) incorporating both uncertainties in these parameters from the D-term calibration and systematic discrepancies between image reconstruction methods.

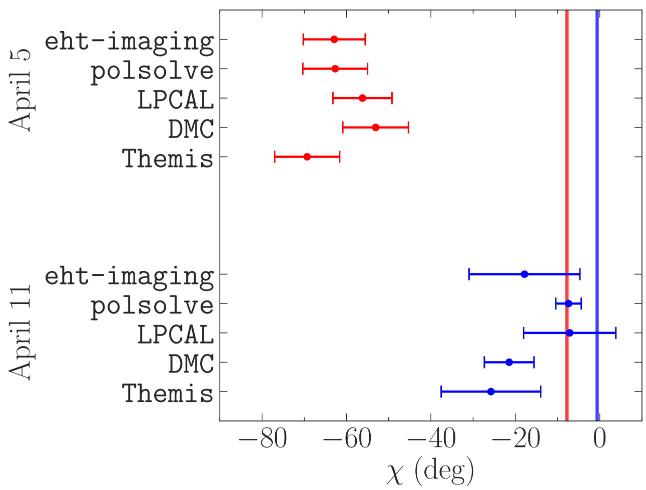

Figure 10. EHT net EVPA integrated within $120\,\mu$as on both April 5 and 11. Vertical lines mark ALMA-only EVPAs and their uncertainties, measured on arcsecond scales on April 5 and 11. (The errorbars of ALMA-only measurements are very small so the shaded bands appear as vertical lines.) The EHT EVPAs mean and standard deviation are estimated from a distribution of EVPAs of 1000 images per method/day. These images were generated in Monte Carlo simulations described in Section 5.2.

The polarimetric image stability analysis shows that the mean total EVPA integrated over the EHT images ranges from $\chi\sim-70^\circ$ to $\chi\sim-55^\circ$ (on April 5) and $\chi\sim-25^\circ$ to $\chi\sim-10^\circ$ (on April 11) depending on the method (Figure 10). The EHT-measured EVPA in the core is significantly offset from the EVPA measured by ALMA on large scales on both analyzed days implying again that the extended component within the central arc-second is polarized as well. We also note that in both EHT and ALMA-only observations, the EVPA swings in the counter-clockwise direction from April 5 to 11.

The EHT 4 GHz bandwidth (two contiguous bands centered at 227.1 and 229.1 GHz) and uncertainties in the D-term calibration do not allow us to precisely quantify the resolved rotation measure (RM) in the EHT image (however see Appendix I for estimates of rotation measure upper limits using low and high band data). The implications of Faraday rotation depend on the magnitude, location, and nature of the Faraday screen. Goddi et al in prep (2020) report contemporaneous ALMA measurements of the RM in M87 ranging from 1.5×10^5 rad m^{-2} to -0.4×10^5 rad m^{-2}. Interpreted as the result of an external Faraday screen, these imply rotations of less than 15°, with day-to-day swings of up to 9° (for the EHT observing frequency, a change of $\Delta{\rm RM}=10^5\,{\rm rad/m^2}$ would cause a swings in the EVPA of $\Delta\chi$=9.7° $(\Delta{\rm RM}/10^5{\rm rad/m^2})$); there is no evidence that such day-to-day swings are present in the reconstructed images.

We measure a larger change in the total EVPA in the EHT images from April 5 to 11, than implied by the variable ALMA RMs, which suggests that the evolution we observe is partially intrinsic to the source rather than an evolution of an external Faraday screen alone. Intrinsic polarimetric evolution is also supported by the changes in Stokes $\mathcal{I}$ alone and by the changes in the distribution of polarized flux density in our polarimetric images (Figure 6 does not show a simple uniform rotation of EVPAs between April 5 and April 11).

Larger RMs remain possible when additional structure is invoked between sub-milliarcsecond and arcsecond scales. A two-component model, comprised of variable compact and static extended emission regions, with associated static Faraday screens (such model is introduced in Goddi et al in prep 2020), is capable of reproducing both the magnitude and interday variability of the ALMA RMs, and suggests that the RM relevant for the images presented here could be of order -5×10^5 rad m^{-2}, corresponding to EVPA rotations of order 50° (Goddi et al in prep 2020). However, the location of this static Faraday screen is presently unconstrained, and may be either external or internal to the emission region (see Appendix I). Starting from 2018, the EHT observes simultaneously in 212.1-216.1 GHz and 226.1-230.1 GHz frequency bands (Paper II). This development should allow us to quantify the resolved RM and address the intrinsic polarimetric variability of M87 with better precision in the future.

Finally, in this letter, we discuss only the linear polarization images. Given its small magnitude, Stokes $\mathcal{V}$ is significantly more sensitive to calibration choices and residual errors than the linear polarization components. For that reason a full analysis of the circular polarization structure in M87 will be presented separately.

7. SUMMARY

We presented polarimetric calibration and polarimetric imaging of the EHT 2017 data on the 230 GHz core of M87 on scales comparable to the supermassive black hole event horizon. Our analysis follows up on the M87 total intensity data calibration, image reconstructions, and model fits presented in Paper III; Paper IV; Paper VI.

We employed multiple distinct methods for polarimetric calibration and polarimetric imaging. All methods were first tested on a suite of synthetic data. When applied to M87, they consistently show that the polarized emission is predominantly from the south-west quadrant. In all reconstructions, the polarization vectors are organized into a similar coherent pattern roughly oriented along the ring. In all reconstructions, both the image-integrated net linear polarization fraction and the average resolved polarization fraction on the ring are consistent to within a few percent. We observed signatures of evolution in the ring's polarization from April 5 to April 11, the full length of the EHT 2017 observing campaign. In this work, we demonstrated that the main polarimetric characteristics of the M87 ring are robust to D-term calibration uncertainties and to the choice of image reconstruction algorithm, though the detailed source structure (particularly in low brightness regions) is still limited by the EHT's very sparse (u, v) coverage and thus depends sensitively on choices made in the image reconstruction and calibration process.

The high-angular-resolution observation with the EHT, on unprecedented scales of $\sim 20\,\mu$as $\approx 2.5R_{\rm Sch}$, allow us for the first time to reconstruct the geometry of magnetic fields in the immediate vicinity of the event horizon of the M87 supermassive black hole. The physical interpretation of our polarimetric images and the full discussion of horizon-scale magnetic field geometries consistent with the EHT images are presented in the accompanying Letter (Paper VIII).

Facilities: EHT, ALMA, APEX, IRAM:30m, JCMT, LMT, SMA, ARO:SMT, SPT .

Software: AIPS (Greisen 2003), ParselTongue (Kettenis et al. 2006), GNU Parallel (Tange 2011), `eht-imaging` (Chael et al. 2016), Difmap (Shepherd 2011), Numpy (van der Walt et al. 2011), Scipy (Jones et al. 2001), Pandas (McKinney 2010), Astropy (The Astropy Collaboration et al. 2013, 2018), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007), THEMIS (Broderick et al. 2020b).

The authors of the present paper 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; an Alfred P. Sloan Research Fellowship; Allegro, the European ALMA Regional Centre node in the Netherlands, the NL astronomy research network NOVA and the astronomy institutes of the University of Amsterdam, Leiden University and Radboud University; the Black Hole Initiative at Harvard University, through a grant (60477) from the John Templeton Foundation; the China Scholarship Council; Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile, via PIA ACT172033, Fondecyt projects 1171506 and 3190878, BASAL AFB-170002, ALMA-conicyt 31140007); Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico, projects U0004-246083, U0004-259839, F0003-272050, M0037-279006, F0003-281692, 104497, 275201, 263356); 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, projects IN112417 and IN112820); the European Research Council Synergy Grant "BlackHoleCam: Imaging the Event Horizon of Black Holes" (grant 610058); the Generalitat Valenciana postdoctoral grant APOSTD/2018/177 and GenT Program (project CIDEGENT/2018/021); MICINN Research Project PID2019-108995GB-C22; 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); Joint Princeton/Flatiron and Joint Columbia/Flatiron Postdoctoral Fellowships, research at the Flatiron Institute is supported by the Simons Foundation; 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, QYZDJSSW- SYS008, ZDBS-LY-SLH011); 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 Malaysian Fundamental Research Grant Scheme (FRGS) FRGS/1/2019/STG02/UM/02/6; 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 and 80NSSC20K1567, NASA Astrophysics Theory Program grant 80NSSC20K0527, and Hubble Fellowship grant HST-HF2-51431.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555,); 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-1312651, AST-1337663, AST-1440254, AST-1555365, AST-1615796, AST-1715061, AST-1716327, AST-1716536, OISE-1743747, AST-1816420, AST-1903847, AST-1935980); the Natural Science Foundation of China (grants 11573051, 11633006, 11650110427, 10625314, 11721303, 11725312, 11933007, 11991052); a fellowship of China Postdoctoral Science Foundation (2020M671266); 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, Basic Research Support Grant 2019R1F1A1059721); 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 Spanish Ministerio de Economía y Competitividad (grants AYA2015-63939-C2-1-P, AYA2016-80889-P, PID2019-108995GB-C21); 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 Consejería de Economía, Conocimiento, Empresas y Universidad of the Junta de Andalucía (grant P18-FR-1769), the Consejo Superior de Investigaciones Científicas (grant 2019AEP112); 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 programme under grant agreement No 730562 RadioNet; ALMA North America Development Fund; the Academia Sinica; Chandra TM6- 17006X; the GenT Program (Generalitat Valenciana) Project CIDEGENT/2018/021. 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. This research was enabled in part by support provided by Compute Ontario (http://computeontario.ca), Calcul Quebec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca). We thank the staff at the participating observatories, correlation centers, and institutions for their enthusiastic support. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2016.1.01154.V. ALMA is a partnership of the European Southern Observatory (ESO; Europe, representing its member states), NSF, and National Institutes of Natural Sciences of Japan, together with National Research Council (Canada), Ministry of Science and Technology (MOST; Taiwan), Academia Sinica Institute of Astronomy and Astrophysics (ASIAA; Taiwan), and Korea Astronomy and Space Science Institute (KASI; Republic of Korea), in cooperation with the Republic of Chile. The Joint

ALMA Observatory is operated by ESO, Associated Universities, Inc. (AUI)/NRAO, and the National Astronomical Observatory of Japan (NAOJ). The NRAO is a facility of the NSF operated under cooperative agreement by AUI. APEX is a collaboration between the Max-Planck-Institut für Radioastronomie (Germany), ESO, and the Onsala Space Observatory (Sweden). The SMA is a joint project between the SAO and ASIAA and is funded by the Smithsonian Institution and the Academia Sinica. The JCMT is operated by the East Asian Observatory on behalf of the NAOJ, ASIAA, and KASI, as well as the Ministry of Finance of China, Chinese Academy of Sciences, and the National Key R&D Program (No. 2017YFA0402700) of China. Additional funding support for the JCMT is provided by the Science and Technologies Facility Council (UK) and participating universities in the UK and Canada. The LMT is a project operated by the Instituto Nacional de Astrófisica, Óptica, y Electrónica (Mexico) and the University of Massachusetts at Amherst (USA). The IRAM 30-m telescope on Pico Veleta, Spain is operated by IRAM and supported by CNRS (Centre National de la Recherche Scientifique, France), MPG (Max-Planck-Gesellschaft, Germany) and IGN (Instituto Geográfico Nacional, Spain). The SMT is operated by the Arizona Radio Observatory, a part of the Steward Observatory of the University of Arizona, with financial support of operations from the State of Arizona and financial support for instrumentation development from the NSF. The SPT is supported by the National Science Foundation through grant PLR- 1248097. Partial support is also provided by the NSF Physics Frontier Center grant PHY-1125897 to the Kavli Institute of Cosmological Physics at the University of Chicago, the Kavli Foundation and the Gordon and Betty Moore Foundation grant GBMF 947. The SPT hydrogen maser was provided on loan from the GLT, courtesy of ASIAA. The EHTC has received generous donations of FPGA chips from Xilinx Inc., under the Xilinx University Program. The EHTC has benefited from technology shared under open-source license by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER). The EHT project is grateful to T4Science and Microsemi for their assistance with Hydrogen Masers. This research has made use of NASA's Astrophysics Data System. We gratefully acknowledge the support provided by the extended staff of the ALMA, both from the inception of the ALMA Phasing Project through the observational campaigns of 2017 and 2018. We would like to thank A. Deller and W. Brisken for EHT-specific support with the use of DiFX. We acknowledge the significance that Maunakea, where the SMA and JCMT EHT stations are located, has for the indigenous Hawaiian people.

APPENDIX

A. POLARIMETRIC DATA ISSUES

In this section we describe station-specific issues and present the results of a set of validation tests and refinements in the calibration that have been performed on the EHT data, prior to the calibration of the instrumental polarization and the final reconstruction of the full-Stokes EHT images.

A.1. *Instrumental polarization of ALMA in VLBI mode*

Phased ALMA records the VLBI signals in a basis of linear polarization, which need a special treatment after the correlation (Martí-Vidal et al. 2016; Matthews et al. 2018). The post-correlation conversion of the ALMA data from linear basis into circular has some implications for the kind of instrumental polarization left after fringe fitting. As discussed in Goddi et al. (2019), any offset in the estimate of the phase difference between the X and Y signals of the ALMA antenna used as the phasing reference (an offset likely related to the presence of a non-zero Stokes V in the polarization calibrator) maps into a post-conversion polarization leakage that can be modelled as a symmetric pure-imaginary D-term matrix (i.e., $D_R = D_L = i\Delta$). The amplitude of the ALMA D-terms, Δ, can be approximated (to a first order) as the value of the phase offset between X and Y in radians (Goddi et al. 2019). Hence, we expect the D_R and D_L estimates for ALMA to be found along the imaginary axis and to be of similar amplitude.

Furthermore, the ALMA feeds in Band 6 (i.e., the frequency band used in the EHT observations) is rotated by 45 degrees with respect to the azimuth axis, which introduces a phase offset between the RCP and LCP post-converted signals that has to be corrected after the fringe fitting. This offset can be applied as a global phase added (subtracted) to the RL and LR correlation products in all baselines (since ALMA has been used as the reference antenna in the construction of the global fringe-fitting solutions). We have applied this 45 degrees rotation to all the visibilities before performing the analysis described in this paper. Hence, the absolute position angles of the electric vectors (EVPA) derived from our EHT observations are properly rotated into the sky frame. This property of the ALMA-VLBI observations (see Appendix D) gives us absolute EVPA values instantaneously.

A.2. *Instrumental polarization of the LMT*

The LMT shows an unexpectedly high leakage signal with a large delay of $\sim$1.5 ns, which affects the cross-polarization phase spectra of the baselines related to the LMT. All LMT baselines show secondary instrumental fringes in the RL and LR correlations, with amplitudes similar to (and even higher than, for the case of sources with low intrinsic polarization) that of the main fringe. These instrumental fringes are minimum in the parallel-hand correlations (RR and LL), but relatively high in the cross-polarization hands and are related to strong polarization leakage likely due to reflections in the optical setup of the LMT receiver used in 2017 (Paper III). For the EHT observations on year 2018 and beyond, the special-purpose interim receiver used at LMT was replaced by a dual-polarization sideband-separating 1.3 mm receiver, with better stability and full 64 Gbps coverage with the rest of the EHT (Paper II), so future polarimetry analyses of the EHT may be free of this instrumental effect from the LMT.

If we take the frequency average over all IFs (the results presented in Papers I to VI are based in this averaging), the effect of this leaked fringe is smeared out, since the average is equivalent to taking the value of the visibility at the peak of the main fringe. This main peak is only affected by the sidelobe of the delayed leaked fringe, with a relative amplitude that we estimate to be of 10–20% of the cross-polarization main fringe. Therefore, the effect of the leaked fringe is small in comparison to the contribution from the ordinary instrumental polarization, which can especially dominate the cross-polarization signal for observations of sources with low polarization like M87, and can be ignored.

A.3. *Instrumental polarization of the SMA*

The dual-polarization observations performed by the SMA use two independent receivers at each antenna to register the RCP and LCP signals. However, the visibility matrices of the baselines related to the SMA are built from the combination of the RCP and LCP streams as if they were registered with one single receiver. Therefore, some of the assumptions made in the RIME (see Equation 6) for the polarimetry calibration (e.g., stable relative phases and amplitude between polarizations) may not apply for the SMA-related visibilities. However, the fringe-fitting of the parallel-hand correlations related to the SMA, as well as the absolute amplitude calibration (both described in Paper III) did account for the drifts in cross-polarization phase and amplitude between the SMA receivers, which makes it possible to model the instrumental polarization using ordinary leakage matrices.

One extra correction that has to be applied to the D-terms of the SMA is a phase rotation between the RCP and LCP leakages, to account for the 45 degrees rotation of the antenna feed with respect to the mount axes. The D-terms shown in Section 4.2 and in Appendix D are corrected by this rotation.

A.4. *Instrumental polarization of the JCMT*

The JCMT was equipped with a single-polarization receiver for these observations, so that only one of the two polarizations can be used at each epoch. Therefore, only one of the two cross-polarization correlations can be computed in all the baselines related to the JCMT; depending on which product is computed, we can only solve for one of the two D-terms of the JCMT (i.e., D_L if RCP is recorded; D_R otherwise). A missing cross-hand correlation in all the baselines to the JCMT implies some limitations for the use of `LPCAL`.

A.5. *Cross-polarization delays*

As it is explained in Martí-Vidal et al. (2016), a byproduct of the use of `polconvert` in VLBI is the calibration of the absolute cross-polarization delays and phases in the stations with polconverted data, which allow for the reconstruction of the absolute EVPAs of the observed sources. The only condition to have this absolute R/L delay and phase calibration is to use the polconverted station (i.e., ALMA, in the case of the EHT) as the reference antenna in the Global Fringe Fitting.

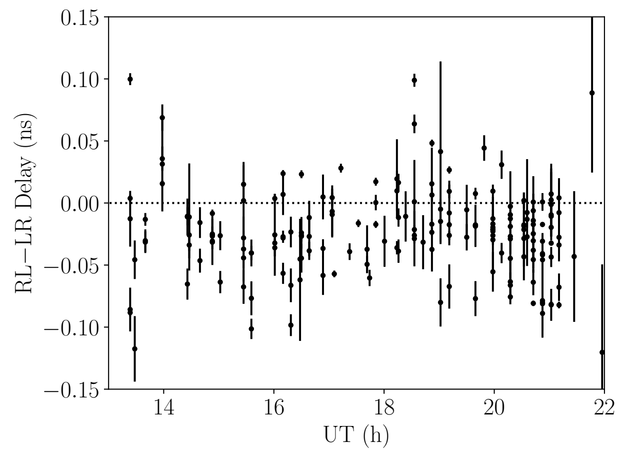

Figure 11. Differences between the RL and LR delays for all baselines and scans with S/N higher than 10 (only scans when ALMA is observing are shown). The level of zero delay is shown as a dashed line.

In Figure 11, we show the difference of multi-band delays between RL and LR after the Global Fringe Fitting (GFF) calibration described in Paper III. All source scans and baselines with an S/N higher than 10 are shown for times when ALMA was participating in the observations. According to Martí-Vidal et al. (2016), the delay difference between RL and LR should be around zero when ALMA is the reference antenna. We see, though, hints of a small global residual delay difference after the GFF calibration (the points are not symmetrically distributed around zero). The weighted average of all the delay differences shown in Figure 11 is $\Delta\tau = -28 \pm 1$ ps. This is a very small delay in absolute value (the amplitude losses due to this delay in each correlation product is lower than 1%), but still detectable at the remarkably high S/N level of the EHT observations.

B. CLOSURE TRACES

B.1. *Definition*

Closure traces (Broderick & Pesce 2020) are calibration-insensitive quantities constructed on station quadrangles from the coherency matrices ρ_{jk} defined in Equation 2:

$$\mathcal{T}_{ijkl} = \frac{1}{2}\mathrm{tr}\left(\rho_{ij}\rho_{kj}^{-1}\rho_{kl}\rho_{il}^{-1}\right). \tag{B1}$$

These data products are a superset of the more familiar closure quantities (closure phases and closure amplitudes), with the additional property that they are independent of instrumental polarization. The closure traces are also independent of any other station-based effects that can be described in a Jones matrix formalism, including the definition of the polarization basis (e.g., the representation of the polarized quantities in terms of linear or circular feeds). The closure traces thus provide a powerful tool with which to make calibration-independent statements regarding polarimetric data and intrinsic source structure.

By analogy with trivial closure phases (see Paper III), trivial closure traces may be constructed on "boomerang" quadrangles, i.e., quadrangles in which a station is effectively repeated in such a way as to make the quadrangle area vanish (Broderick & Pesce 2020). Given two co-located stations i and i', the closure trace $\mathcal{T}_{iji'k}$ reduces to unity.

Each quadrangle $ijkl$ has an associated "conjugate" quadrangle $ilkj$, constructed by reordering the baselines within the coherency matrix product.[8] Conjugate closure trace products can be expressed as

$$\begin{aligned}\mathcal{C}_{ijkl} \equiv \mathcal{T}_{ijkl}\mathcal{T}_{ilkj} = 1 &+ \left(\breve{q}_{ij} - \breve{q}_{kj} + \breve{q}_{kl} - \breve{q}_{il}\right)^2 \\ &+ \left(\breve{u}_{ij} - \breve{u}_{kj} + \breve{u}_{kl} - \breve{u}_{il}\right)^2 \\ &+ \left(\breve{v}_{ij} - \breve{v}_{kj} + \breve{v}_{kl} - \breve{v}_{il}\right)^2 \\ &+ \mathcal{O}\left(\breve{q}^3, \breve{u}^3, \breve{v}^3\right),\end{aligned} \tag{B2}$$

where $\breve{q}_{ij} \equiv \tilde{\mathcal{Q}}_{ij}/\tilde{I}_{ij}$, $\breve{u}_{ij} \equiv \tilde{\mathcal{U}}_{ij}/\tilde{I}_{ij}$, and $\breve{v}_{ij} \equiv \tilde{\mathcal{V}}_{ij}/\tilde{I}_{ij}$. The $\mathcal{C}_{ijkl}$ are identically unity in the absence of intrinsic source polarization and for point sources. Deviations from unity require non-constant interferometric polarization fractions on baselines in the quadrangle $ijkl$, and therefore closure trace products are a robust indicator of polarized source structures (Broderick & Pesce 2020).

B.2. *Implications for polarimetric data quality*

For EHT observations, boomerang quadrangles are formed using the redundant baselines presented by ALMA and APEX[9]. In the upper panel of Figure 12, the phases of all of the trivial closure traces are shown for M87 (blue) and the calibrators (J1924–2914, NRAO 530, 3C 279; grey), constructed from scan-averaged visibility data. The values of these phases are clustered about zero, consistent with the expectation that the trivial closure traces are unity.

The distribution of the normalized residuals provide a direct assessment of the systematic error budget of the polarimetric data independent of the gain and leakage calibration. These residuals are shown in the lower

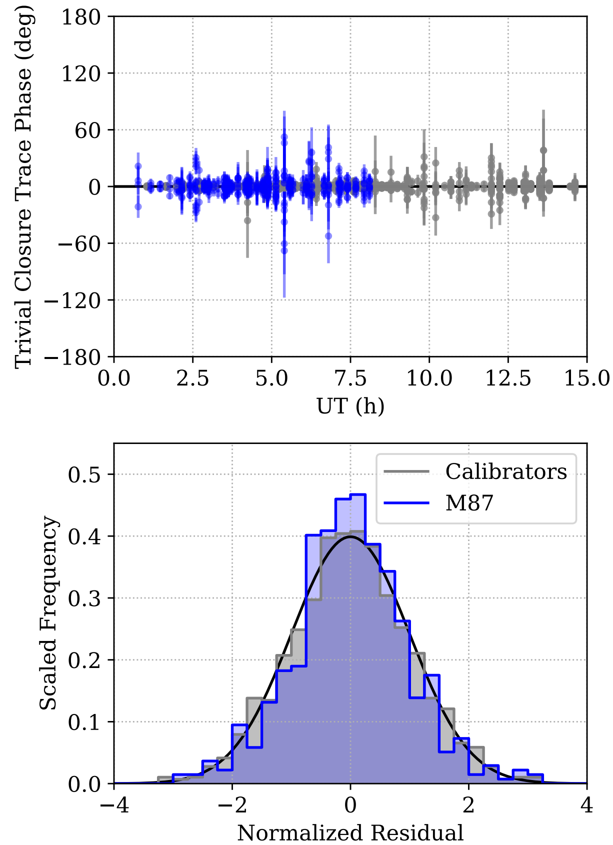

Figure 12. Top: phases of "boomerang" closure traces for M87 (blue) and calibrators (J1924–2914, NRAO 530, 3C 279; grey), i.e., those with a repeated station (here ALMA/APEX) and thus expected to trivially vanish. Bottom: normalized residuals of the trivial closure traces on M87 and the calibrator sources in comparison to a unit-variance normal distribution. In both panels, high- and low-band values are shown for scan-averaged data. We see that these boomerang closure trace phases exhibit the expected clustering around zero.

[8] This conjugate quadrangle is identical to the degenerate quadrangle formed by inverting the numerator and denominator in a standard closure amplitude.

[9] The redundant baselines to SMA and JCMT cannot be used as a result of the single-polarization observations at the latter.

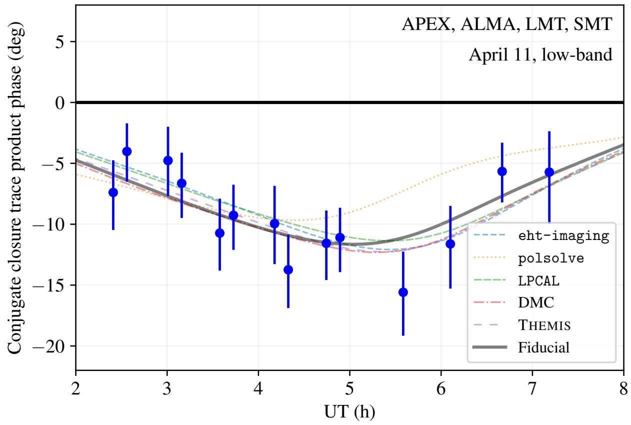

Figure 13. Phase of the conjugate closure trace product constructed from the April 11, 2017 low-band observations on the APEX–ALMA–LMT–SMT and APEX–SMT–LMT–ALMA quadrangles. This quantity is sensitive solely to polarization structure; deviations from zero indicate the presence of non-trivial polarization structure (i.e., a polarization fraction that is not constant across the source; Broderick & Pesce 2020). The colored lines show the same conjugate closure trace product phase for each of the image reconstructions (see Figure 6), and the dark gray solid line shows the same for the fiducial image (see Figure 7).

panel of Figure 12 for both M87 and calibrators. We find that the data match the anticipated unit-variance Gaussian, consistent with an absence of unidentified systematic uncertainties in the polarimetric data.

B.3. *Calibration-insensitive detection of polarization*

Figure 13 shows the phase of the conjugate closure trace product, $\mathcal{C}_{ijkl}$, for a quadrangle pair ALMA–APEX–LMT–SMT and ALMA–SMT–LMT–APEX. The presence of non-zero $\mathcal{C}_{ijkl}$ is a calibration-insensitive indicator of significant polarized structures in the Stokes map. Because the uncertainties of the closure traces on conjugate quadrangles are correlated, the resulting uncertainty in the conjugate closure trace product is typically smaller than would be estimated from assuming independent uncertainties in the individual closure traces. The errors shown have been estimated using Monte Carlo sampling of the constituent visibilities.

B.4. *Calibration-insensitive detection of evolving source structure*

Closure trace phases are shown on a handful of non-trivial quadrangles in Figure 14 for M87. These phases are clearly non-zero and exhibit variations throughout the observing night, consistent with non-trivial source structure. The behavior of the closure trace evolution is similar across neighboring observation days (e.g., April 5/6, April 10/11) and consistent between quadrangles constructed using ALMA (filled markers) and APEX (open markers). The behavior of the closure trace evolution is dissimilar between the April 5/6 and April 10/11 observations, providing direct evidence for an evolving source structure in M87. Because ALMA, LMT, and SMT are nearly co-linear as seen from M87 for much of the observations, and because the closure traces are presumably tracing primarily the Stokes $\mathcal{I}$ emission, the closure trace phases in Figure 14 are very similar to the Stokes $\mathcal{I}$ closure phases shown in Figure 14 of Paper III.

B.5. *Calibration-insensitive probe of evolving polarimetric source structure*

Conjugate closure trace product phases are shown in Figure 15 for each observation day for the ALMA–PV–LMT–SMT and ALMA–SMT–LMT–PV quadrangle pair. There is the appearance of temporal evolution from April 5/6 to April 10/11, with an attendant implication for an evolution in the polarization map of M87 between those periods. However, the paucity of quadrangles exhibiting significant evolution renders this conclusion suggestive at best.

C. DETAILED DESCRIPTION OF ALGORITHMS FOR THE CALIBRATION OF INSTRUMENTAL POLARIZATION

C.1. *Polarimetric Imaging via sub-component fitting: `polsolve`, LPCAL, and GPCAL*

In the sub-component fitting method polarimetric modeling, the instrumental polarization (i.e., the complex D-terms) and the source polarized brightness distribution are estimated simultaneously from the interferometric observables and a fixed estimate of the source brightness distribution $\mathcal{I}(\mathbf{x})$. The sub-component fitting calibration algorithms estimate the D-terms in Equation 4 by modelling the polarized source structure as a disjoint set of N "polarization sub-components", $\mathcal{I}_i(\mathbf{x})$, such that:

$$\mathcal{I}(\mathbf{x}) = \sum_i^N \mathcal{I}_i(\mathbf{x}). \tag{C3}$$

The fractional polarization of each sub-component is assumed to be constant, implying that $\mathcal{Q}(\mathbf{x}) = \sum_i^N q_i\, \mathcal{I}_i(\mathbf{x})$ and $\mathcal{U}(\mathbf{x}) = \sum_i^N u_i\, \mathcal{I}_i(\mathbf{x})$, where q_i and u_i are real-valued constants. In the sub-component fitting method, we therefore assume that the polarized brightness is exactly proportional to $\mathcal{I}_i$ for each source sub-component. This condition is known as the "similarity approximation" and may produce inaccurate estimates of the instrumental polarization for cases of strongly polarized and resolved calibrators (e.g., Cotton 1993) and/or if the sub-division of $\mathcal{I}$ into sub-components is not performed properly. Discussions about the self-similarity assumption can be found in Appendix K.

The `polsolve`, `LPCAL`, and `GPCAL` algorithms determine which values of q_i, u_i, D_R and D_L minimize the difference between the calibrated visibility matrix and

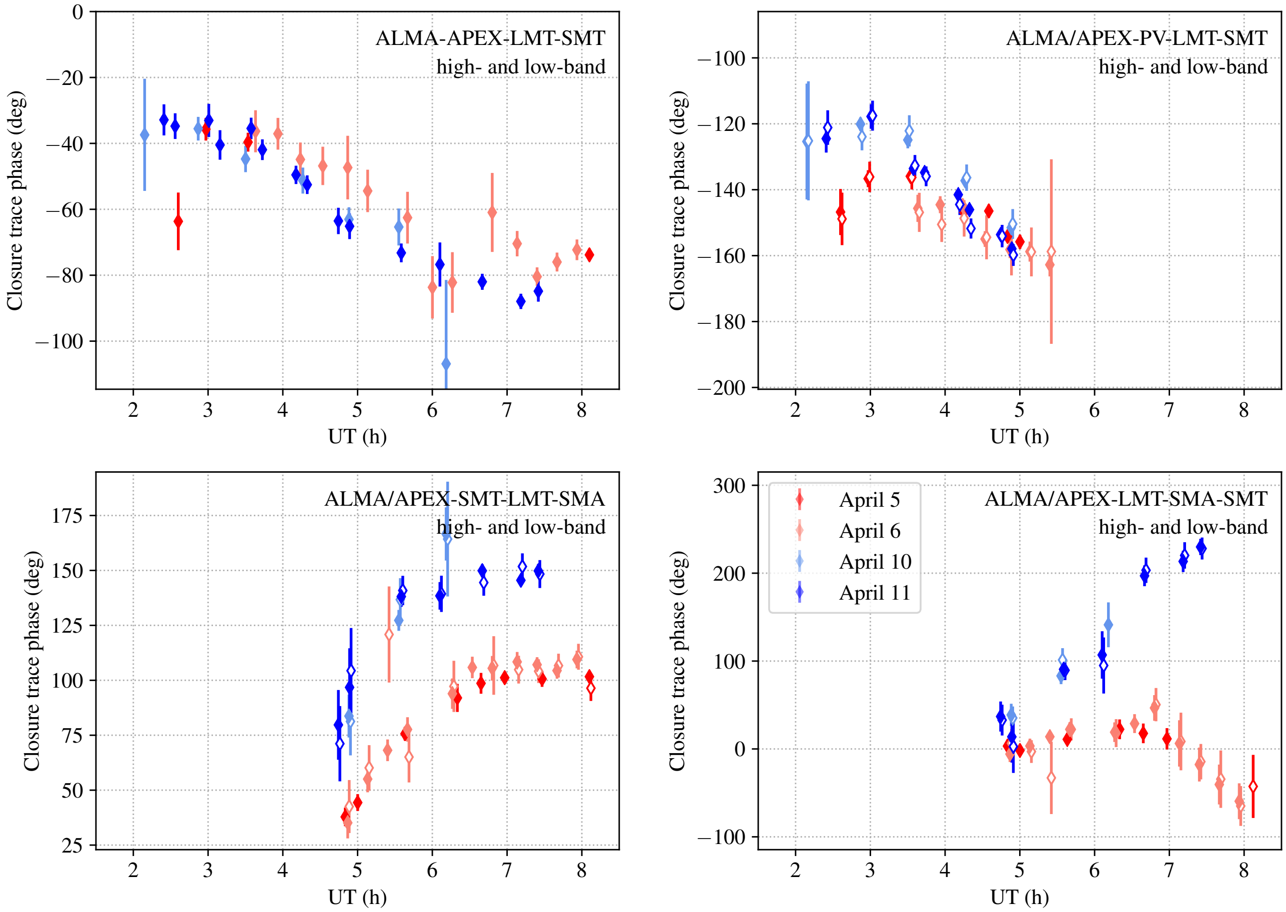

Figure 14. Phases of M87 closure traces on four illustrative quadrangles for the different observation days. Closure traces constructed with ALMA and APEX are shown by filled and open points, respectively. The plotted closure traces represent an average across high- and low-band data. An $S/N > 1$ selection in the final closure trace phase has been applied.

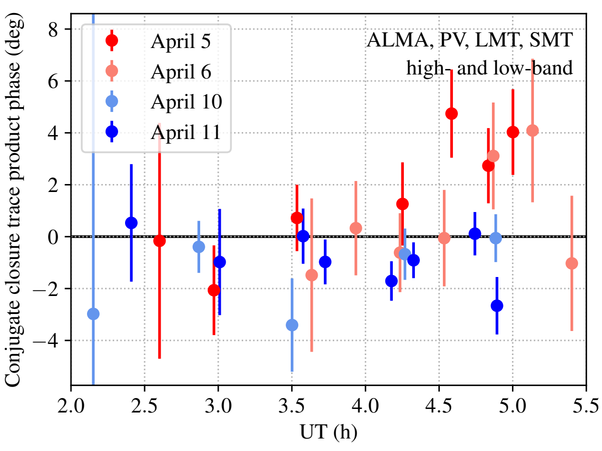

Figure 15. Phase of conjugate closure trace products constructed on the ALMA–PV–LMT–SMT and ALMA–SMT–LMT–PV quadrangles, averaged across high- and low-bands for each of the days on which M87 was observed.

the Fourier-transformed model brightness matrix (Equations 6 and 4). The total number of parameters used in this fit is equal to two times the the number of source sub-components (i.e., $2N$, which correspond to q_i and u_i in $\mathcal{Q}(\mathbf{x}) = \sum_i^N q_i\,\mathcal{I}_i(\mathbf{x})$ and $\mathcal{U}(\mathbf{x}) = \sum_i^N u_i\,\mathcal{I}_i(\mathbf{x})$) plus four times the number of antennas (i.e., $4N_a$, accounting for the real and imaginary parts of the D_R and D_L of each antenna). The error function (or log-likelihood) to be minimized is the sum of the χ^2 values computed for the cross-polarization matrix elements of the RIME, i.e.,

$$\chi^2 = \sum_m^{N_v} w_m \left| RL^c_{kl,m} - \left(\tilde{\mathcal{Q}} + i\tilde{\mathcal{U}}\right)_m \right|^2 + \sum_m^{N_v} w_m \left| LR^c_{kl,m} - \left(\tilde{\mathcal{Q}} - i\tilde{\mathcal{U}}\right)_m \right|^2, \qquad \text{(C4)}$$

where w_m is the weight of the m-th visibility, the index c stands for calibrated visibilities (corrected both for station gains and for instrumental polarization using the current estimate of the D-terms) and N_v is the number of visibilities.

The calibrated visibilities, $RL^c_{kl,m}$ and $LR^c_{kl,m}$, depend on D^k_R, D^k_L, D^l_R and D^l_L (Equation 4), whereas $\tilde{\mathcal{Q}}$

and $\tilde{\mathcal{U}}$ depend on q_i and u_i. The χ^2 minimization solves for the intrinsic source Stokes parameters and the instrumental polarization simultaneously. We note that the effects of instrumental polarization are constant in the frame of the antenna feed for Cassegrain-mounted feeds, whereas the intrinsic source polarization is defined in the sky frame; as a consequence, the changing feed angle of each antenna across the observations (i.e., the Earth rotation during the extent of the observations) allows the model fitting to decouple the antenna D-terms from the Stokes parameters of the source sub-components. For Nasmyth-mounted feeds, there is an additional rotation between cross-polarization introduced by the feed/optics and the telescope itself. We assume a minimal contribution from the antennas themselves, a reasonable assumption for these on-axis telescopes. Equation C4 **implies** that there are several implicit assumptions in the polarimetric modelling of `polsolve` and `LPCAL`. On the one hand, RL^c and LR^c are computed by setting $\mathcal{V} = 0$ (i.e., any circular polarization in the calibrators is neglected, compared to Stokes $\mathcal{I}$). On the other hand, the real and the imaginary parts of the residual visibilities (i.e., either RL^c or LR^c minus the Fourier transforms of the corresponding model brightness distributions) are assumed to be statistically independent.

If the linear polarization structures of calibrators are not similar to their total intensity structures, the breakdown of the similarity approximation can occur. This can be a source of uncertainties in D-term estimation. In this case, one can (nearly) overcome the limitations of the similarity approximation as follows.

1. Obtain the D-terms using the similarity approximation and correct them.
2. Reconstruct Stokes $\mathcal{Q}$ and $\mathcal{U}$ images with CLEAN using the D-term corrected data. For these observations, a pixel size of $1\,\mu$as is small enough for the CLEANing. Compute the model visibilities $\tilde{\mathcal{Q}}$ and $\tilde{\mathcal{U}}$ from the CLEAN models.
3. Solve Equation C4 for D-terms only using the model visibilities constrained in step 2.
4. Iterate over steps 2 and 3 until the fitting solutions and statistics are converged.

Since this scheme is very similar to self-calibration of parallel-hand data, it was named "instrumental polarization self-calibration" (Cotton 1995, Section 15.4.3). This approach would be particularly useful for D-term estimation from calibrators having relatively high fractional polarization and complex linearly polarized structures. The instrumental polarization self-calibration mode is implemented in `GPCAL` and in `polsolve`.

C.2. *Polarimetric Imaging via Regularized Maximum Likelihood: `eht-imaging`*

The package `eht-imaging` (Chael et al. 2016, 2018) implements the image reconstruction via the Regularized Maximum Likelihood (RML). `eht-imaging` solves for an image $\mathbf{X}$ by minimizing an objective function via gradient descent. The objective function $J(\mathbf{X})$ is a weighted sum of data-consistency log-likelihood terms and *regularizer* terms that favor or penalize certain image features. That is, to find an image (in either total intensity, polarization, or both) we minimize

$$J(\mathbf{X}) = \sum_{\text{data terms}\, i} \alpha_i \chi_i^2(\mathbf{X}) + \sum_{\text{regularizers}\, j} \beta_j S_j(\mathbf{X}). \quad \text{(C5)}$$

Picking optimal values of the "hyperparameter" weights α_i and β_j in Equation C5 is an essential task in RML imaging. Here we describe the data terms and regularizers we use for polarimetric imaging, and in Appendix G we describe our method for determining the hyperparameters using parameter surveys.

For polarized image reconstructions, we follow the method laid out in Chael et al. (2016), with the addition of iterative self-calibration of any uncorrected station D-terms. First, we reconstruct a Stokes $\mathcal{I}$ image. For M87, we use the fiducial imaging script for `eht-imaging` developed in Paper IV. In this stage, we iteratively self-calibrate the station amplitude and phase gains to ensure the final gain calibration solution matches our total-intensity image as much as possible. We then solve for the linearly polarized brightness on top of the fixed Stokes $\mathcal{I}$ image. In between rounds of polarimetric imaging, we iteratively solve for the D-terms by minimizing the χ^2 between the real (gain-calibrated) data and synthetic data from the current image reconstruction corrupted with Jones matrices (Equation 4). We do not use any linearized approximations of the effects of the Jones matrices when solving for the D-terms, but throughout we assume the model image has no circular polarization ($\mathcal{V} = 0$). The `eht-imaging` pipeline thus alternates between rounds of polarimetric imaging and D-term calibration; often it takes many successive rounds ($n_{\rm iter} \approx 50-100$) for the process to converge on a stable D-term solution.

The data used as input to the `eht-imaging` reconstructions has had the overall time-dependent station amplitude and phase gains calibrated using the SMILI fiducial image from Paper IV, and the ALMA, APEX, SMA, and JCMT D-terms have been corrected using the zero-baseline solutions described in Section 4.2. The data are scan-averaged.

We first reconstruct the Stokes $\mathcal{I}$ using the same fiducial imaging script for `eht-imaging` developed in Paper IV. We fix the image field of view at $120\,\mu$as and solve for a grid of 64×64 pixels. In the Stokes $\mathcal{I}$ imaging, the total flux density is constrained to be 0.6 Jy. We self-calibrate the station amplitude and phase gains (assuming $G_R = G_L$) to our final Stokes I image. Having extensively explored the imaging parameter space for Stokes $\mathcal{I}$ imaging in Paper IV, we do not vary these parameters in our polarimetric imaging surveys. After selfcalibrating to our final total intensity image, we drop zero-baselines for the polarimetric imaging stage.

In defining an objective function of the form in Equation C5 for the *polarized* image reconstruction, we consider two log-likelihood χ^2 terms; one computed using the RL* polarimetric visibility $\tilde{p} = \tilde{\mathcal{Q}} + i\tilde{\mathcal{U}}$, and one using the visibility domain polarimetric ratio $\breve{m} = \tilde{p}/\tilde{\mathcal{I}}$. $\chi^2_{\breve{m}}$ is immune to any residual station gain error left over from Stokes $\mathcal{I}$ imaging, while $\chi^2_{\tilde{p}}$ is not. We use two regularizers on the polarized flux. First, the Holdaway-Wardle (Holdaway & Wardle 1990) regularizer S_{HW} (Equation 13 of Chael et al. 2016) acts like an entropy term that prefers image pixels take a value less than $m_{\max} = 0.75$. This regularizer encourages image pixels to stay below the theoretical maximum polarization fraction for synchrotron radiation, but it is not a hard limit. Second, the total variation (TV) regularizer S_{TV} (Rudin et al. 1992) acts to minimize in both the real and imaginary part of the complex polarization brightness distribution from a pixel-to-pixel distribution (Equation 15 of Chael et al. 2016).

Taken together, the objective function we minimize in polarimetric imaging is

$$J_{\rm pol}(\mathcal{Q},\mathcal{U}) = \alpha_p \chi^2_{\tilde{p}} = \alpha_m \chi^2_{\breve{m}} - \beta_{HW} S_{HW} - \beta_{TV} S_{TV}. \tag{C6}$$

The relative weighting between the data constraints and the regularizer terms is set by the four hyperparameters $\alpha_P, \alpha_m, \beta_{HW}$,and β_{TV}.

We solve for the polarized flux distribution that minimizes Equation C6 parameterized by the fractional polarization m and EVPA ξ in each pixel. The Stokes $\mathcal{I}$ image is fixed in the polarimetric imaging step and defines the region where polarimetric flux is allowed. To ensure our solution respects $Q^2 + U^2 < I^2$ everywhere, we transform the fractional polarization m in each pixel from the range $m \in [0, 1]$ to $\kappa \in (-\infty, \infty)$ and solve for κ (See Appendix D of Chael et al. 2016). In the `eht-imaging` script for EHT M87 observations, we solve for the pixel values of κ and ξ that minimize the objective function by gradient descent, and we then transform $(\mathcal{I}, \kappa, \xi) \rightarrow (\mathcal{I}, \mathcal{Q}, \mathcal{U})$. We often restart the gradient descent process several times, using the output of the previous round of imaging blurred by a $20\,\mu$as Gaussian kernel as the new initial point.

C.3. *Polarimetric imaging as posterior exploration: DMC and* Themis

In this section we describe two Markov chain Monte Carlo (MCMC) schemes developed for polarimetric imaging. Both MCMC codes model the polarized emission structure on a Cartesian grid of intensity points, with the Stokes vector parameterized using a spherical (Poincaré) representation,

$$\begin{pmatrix} \mathcal{I}_i \\ \mathcal{Q}_i \\ \mathcal{U}_i \\ \mathcal{V}_i \end{pmatrix} = \mathcal{I}_i \begin{pmatrix} 1 \\ \ell_i \cos(\xi_i)\sin(\varsigma_i) \\ \ell_i \sin(\xi_i)\sin(\varsigma_i) \\ \ell_i \cos(\varsigma_i) \end{pmatrix}, \tag{C7}$$

where the index i runs over individual grid points. Stokes visibilities are generated from the gridded emission structure via a direct Fourier transform (i.e., treating each grid point as a point source), and the visibilities are then multiplied with a smoothing kernel to impose image continuity. The parallel- and cross-hand visibilities on each baseline are then computed from the Stokes visibilities using Equation 7, and the gains and leakage terms are applied to the model visibilities using a Jones matrix formalism (see Equation 6). The model and data visibilities are ultimately compared via complex Gaussian likelihood functions for each of the parallel- and cross-hand data products independently, with the total likelihood taken to be the product of the individual likelihoods for all parallel- and cross-hand data products.

C.3.1. *DMC*

We introduce a new D-term Modeling Code (DMC) that utilizes the Hamiltonian Monte Carlo (HMC) sampler implemented in the PyMC3 probabilistic programming Python package (Salvatier et al. 2016) to perform posterior exploration. We briefly describe the relevant aspects of the DMC analysis in this section; a more thorough description of the software will be provided in a forthcoming publication (D. W. Pesce et al. *in prep.*). Prior to fitting, we coherently average the visibility data on a per-scan basis and flag the intrasite baselines.

Within the DMC framework, the $\mathcal{I}$ image axes are aligned with the equatorial coordinate axes. The pixel intensities are constrained to sum to a total flux via the imposition of a flat Dirichlet prior, and the total flux parameter is restricted to be positive via a uniform prior on the range (0,2) Jy. The radial Stokes parameter (ℓ_i in Equation C7) is sampled from a unit uniform prior, and the angular Stokes parameters (ξ_i and ς_i in Equation C7) are uniformly sampled on the sphere. We multiply the model visibilities by a circular Gaussian kernel to impose image smoothness.

In DMC, both the right- and left-hand complex station gains are modeled independently on every scan, save for a single reference station (chosen to be ALMA) that is constrained to have zero right- and left-hand gain phase at all times. We impose log-normal priors on the gain amplitudes and wrapped uniform priors[10]

[10] A "wrapped" or "circular" uniform distribution is defined on the unit circle and has constant probability density for all angles. That is, an angular variable drawn from a wrapped uniform distribution is being sampled uniformly on the unit circle.

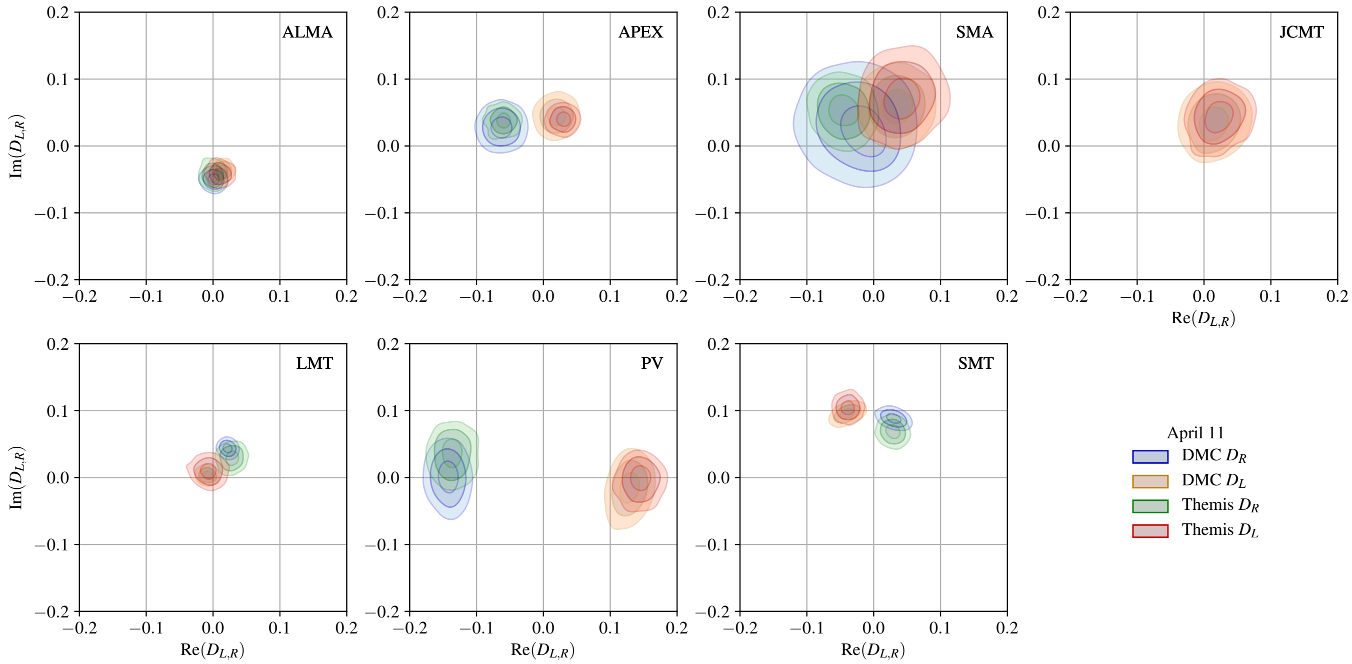

Figure 16. Leakage posteriors for individual stations from DMC and Themis reconstructions of M87 on April 11. Because JCMT only records a single polarization, only lefthand D-terms are shown. The plotted contours enclose 50%, 90%, and 99% of the posterior probability, and show a large degree of overlap for all stations despite considerable differences in the underlying model specifications.

on the gain phases. The right- and left-hand leakage amplitudes are sampled from a unit uniform prior, and the leakage phases are sampled from a wrapped uniform prior.

The DMC likelihood variances are set to the quadrature sum of the data thermal variances and a systematic component that is modeled as the square of a fraction of the Stokes $\mathcal{I}$ visibility amplitude; this fractional uncertainty parameter is sampled from a unit uniform prior.

C.3.2. Themis

The existing imaging method decribed in Broderick et al. (2020a) has been extended to polarization reconstructions. This makes use of a deterministic even-odd swap tempering scheme (Syed et al. 2019) using the HMC sampling kernel from the Stan package (Carpenter et al. 2017). Here we briefly summarize the implementation and assumptions underlying the Themis polarization map reconstructions; more detail on these points will be presented elsewhere (A. E. Broderick et al. *in prep.*).

As with DMC, all Themis analyses are performed on coherently scan-averaged visibility data. Unlike the DMC analysis, intrasite baselines are included to facilitate gain and leakage calibration. This is enabled by the inclusion of a large, uniformly polarized Gaussian to model the milliarcsecond-scale structure (see, e.g., Broderick et al. 2020a; Paper IV)

Themis models the polarized image as a small number of control points located on a rectilinear grid, from which the fields $\mathcal{I}$, ℓ, ξ, and $\cos(\varsigma)$ are constructed via an approximate cubic spline in a fashion similar to Broderick et al. (2020a). The field of view and orienation of the rectilinear grid are fit parameters and permitted to vary. In this way the effective resolution is reconstructed from the data itself. Logarithmic priors are adopted on $\mathcal{I}$ and ℓ, flat priors are adopted on ξ and $\cos(\varsigma)$ with the natural limits.

Complex station gains are reconstructed via the Laplace approximation (see Section 6.8 of Broderick et al. 2020b). The right- and left-hand complex station gains are constrained to be equal, and permitted to vary independently on every scan. Log-normal priors are imposed on the station gain amplitudes. The real and imaginary components of the right- and left-hand leakages are treated as additional model parameters, with each component sampled uniformly on [-1,1].

Unless otherwise indicated, Themis analyses shown here used a 5×5 raster grid, consistent with that typically necessary to capture features on the scale of the EHT beam within the field of view imposed by the shortest intersite baselines. A 3% systematic noise component was added in quadrature to the thermal uncertainties to capture non-closing errors in the scan-averaged visibilities. These are similar to the magnitude of fractional systematic error inferred from the DMC analyses.

Table 3. Daily average D-terms for ALMA derived via the multi-source intra-site method.

Date	Band	D_R (%)	D_L (%)
Apr 5	low	$0.30 - 2.80i$ (± 0.70)	$-1.42 - 3.74i$ (± 0.70)
	high	$-0.17 - 4.10i$ (± 0.60)	$-1.09 - 4.02i$ (± 0.60)
Apr 6	low	$0.60 - 5.45i$ (± 0.40)	$-0.53 - 6.08i$ (± 0.40)
	high	$-0.09 - 1.52i$ (± 0.30)	$-0.75 - 1.66i$ (± 0.30)
Apr 7	low	$1.12 - 7.10i$ (± 0.70)	$-0.46 - 5.77i$ (± 0.70)
	high	$1.25 - 4.93i$ (± 0.70)	$-0.37 - 4.00i$ (± 0.70)
Apr 10	low	$0.78 - 2.61i$ (± 0.30)	$-0.40 - 2.82i$ (± 0.30)
	high	$-0.02 - 3.04i$ (± 0.30)	$-0.56 - 3.92i$ (± 0.30)
Apr 11	low	$-0.15 - 6.33i$ (± 0.50)	$-0.80 - 6.09i$ (± 0.50)
	high	$-0.29 - 5.19i$ (± 0.40)	$-0.76 - 5.07i$ (± 0.40)

Note – The D-term posterior distributions are assumed to be circular Gaussians in the complex plane.

Table 4. Campaign-average D-terms for APEX, JCMT and SMA derived via the multi-source intra-site method.

Station	D_R (%)	D_L (%)
APEX	$-8.67 + 2.96i$ (± 0.70)	$4.66 + 4.58i$ (± 1.20)
JCMT	$-0.09 - 2.29i$ (± 1.80)	$-0.46 + 3.34i$ (± 0.60)
SMA	$-1.73 + 4.81i$ (± 1.00)	$2.79 + 4.00i$ (± 2.20)

C.3.3. Themis-*DMC Leakage Posterior Comparison*

In Figure 16 we show a comparison between the leakage posteriors for all stations, as determined by DMC and Themis fits to the April 11 observations of M87. Despite the various different assumptions and model specifications, we find excellent agreement in both the means and shapes of the posterior distributions recovered from both methods. The modest discrepancies between the posteriors shown in Figure 16 are associated with the different treatment of systematic uncertainty and right/left gain ratios between the two methods; when these model choices are homogenized, the DMC and Themis fits to both synthetic data sets and to the M87 data return indistinguishable posteriors. Notably, both model treatments of the Stokes map appear to be comparably capable of capturing the source structure.

D. INTRA-SITE D-TERM VALIDATION

We present the final D-terms for ALMA in Table 3 and for APEX, JCMT and SMA in Table 4.

JCMT can only record one of two polarization channels at a given time, see Appendix A.4. Therefore, the coherency matrix given in Equation 2 is incomplete for the JCMT–SMA baseline; the missing cross-polarization components on all baselines to the JCMT imply that the relation between visibilities and polarized brightness distribution is an under-determined problem. Fortunately, when fitted in combination with the ALMA–APEX baseline, the Stokes parameters of the unresolved source are determined by the latter. This information is used simultaneously to fit for the JCMT and SMA D-terms. In this fit, only the D-terms affecting the observed cross-polarization product can be estimated, which means that, for each JCMT polarization configuration, only one of the two D-terms of each station (SMA and JCMT) can be determined.

These D-terms, being instrumental properties, are expected to remain constant across observations of different target sources and observations carried out across multiple days. In the case of ALMA, however, D-terms are generated due to an offset in the relative phase calibration between the X and Y linear polarization of the reference ALMA antenna in the VLBI phasing procedure (Martí-Vidal et al. 2016; Matthews et al. 2018; Goddi et al. 2019). The estimated relative X-Y phase at ALMA may change between epochs due to, e.g., a change in the reference antenna[11], a re-setting of the ALMA delay calibration, or the use of different calibrators in the polarization calibration process. Therefore, day-to-day variations in ALMA D-terms are expected. D-terms are expected to have a frequency dependence for all stations, hence we obtain separate estimates for each 2 GHz band.

The D-terms fitted for ALMA are dominated by an imaginary component and indicate day-to-day variation along the imaginary axis, as expected from the physical understanding of the leakage origin (Martí-Vidal et al. 2016; Matthews et al. 2018; Goddi et al. 2019). The dispersion in D-term estimates between days and bands is remarkably low for APEX. The fitting is thus consistent among days as expected: the APEX hardware appears stable across the whole EHT campaign. Similar to APEX, SMA and JCMT should not have varying D-terms across epochs. We therefore derive campaign-average D-terms for these three stations from the day-by-day estimates, combining bands. For ALMA, per-day/band D-term estimates are used.

We validate our D-term calibration via intra-site baseline properties using three methods: comparing intra-site baseline source properties to interferometric-ALMA measurements; comparing SMA intra-site leakage estimates to interferometric-SMA estimates; and comparing `polsolve` leakage estimation to point-source polarimetric modeling with the `eht-imaging` library. We additionally motivate leakage calibration using band-

[11] In case the reference antenna is changed within one epoch, the APP scripts can re-reference the polarizer phases to any other antenna, though with some loss of precision (Goddi et al. 2019).

averaged products from intra-band leakage studies for ALMA–APEX.

Simultaneously to our VLBI observations, ALMA also observes as an interferometric array (referred to as ALMA-only) in a linear-polarization basis. This array data is used for ALMA-VLBI calibration in the Quality Assurance process at ALMA (QA2; Goddi et al. 2019), and provides source-integrated information for calibration refinement and validation, such as total flux densities or polarization properties. Given that our intra-site baselines do not resolve the observed sources, the source-integrated properties from ALMA–APEX, SMA–JCMT and the core component of ALMA-only should match. We show our validation of the derived source polarimetric properties from the intra-site D-term fitting against QA2 ALMA-only estimates in the left (for $\mathcal{Q}$) and center (for $\mathcal{U}$) panels of Figure 17. There is a strong correlation between the Stokes parameters of all sources derived from the ALMA-only observations (Goddi et al. 2019) and the estimates from the ALMA–APEX intra-site VLBI baseline. This correlation can be seen as a further validation test of the quality of the EHT polarimetric calibration.

The polarimetric leakage of the SMA is well characterized, with D-terms of only a few percent expected for observations near the 233.0 GHz tuned frequency of the quarter-wave plates (Marrone 2006; Marrone & Rao 2008). In addition to historical measurements of leakage, near-in-time polarimetric observations of sources with the SMA also allowed us to compute quasi-simultaneous leakage estimates that can be compared with our intra-site method estimates. Observations of 3C 454.3, M87, and 3C 279 within a month of our EHT campaign provided an upper limit of 10% for D-terms, consistent with our intra-site method estimates, and stable across days at the 1% level.

We also validated the `polsolve` leakage estimates using point-source modeling within the `eht-imaging` and DMC libraries. Both modeling schemes assume a constant polarization fraction and EVPA for the point source. The DMC model fits to cross-hand and parallel-hand visibilities by incorporating right- and left-hand station gains as model parameters, while the `eht-imaging` model fits to gain-independent "polarimetric closure" data products consisting of the ratio of cross-hand visibilities to parallel-hand visibilities on a single baseline (see, e.g., Blackburn et al. 2020),

$$\mathcal{K}_{jk} = \frac{R_j L_k^* \times L_j R_k^*}{R_j R_k^* \times L_j L_k^*}. \tag{D8}$$

Both models use Gaussian likelihood functions for their respective data products.

In Figure 18 we compare the multi-source `polsolve` leakage estimates to multi-source `eht-imaging` fits and single-source (using 3C 279) DMC fits; both `eht-imaging` and DMC have fit only to the ALMA-APEX baseline, while the `polsolve` estimates additionally fit to the SMA-JCMT baseline. We find that the leakage terms recovered by all three methods are consistent with one another, with an uncertainty-weighted mean absolute deviation across all days and bands of <1% in absolute leakage between any two methods.

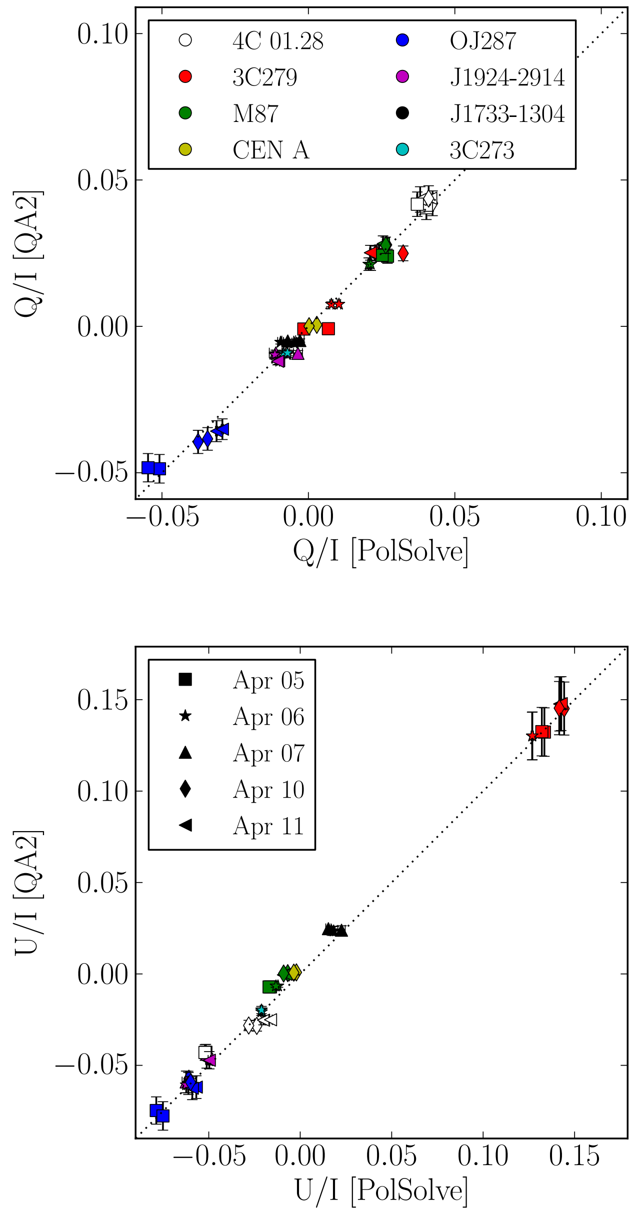

Figure 17. Comparison of source-integrated Stokes $\mathcal{Q}$ and $\mathcal{U}$ estimates from intra-site EHT baselines using the multi-source fitting mode of `polsolve` to those from ALMA-only observations (Goddi et al. 2019).

Furthermore, our leakage estimation methods used band-averaged data products. Given the high S/N of the detections between ALMA and APEX, there are strong detections in all four correlation products (i.e., RR, RL,

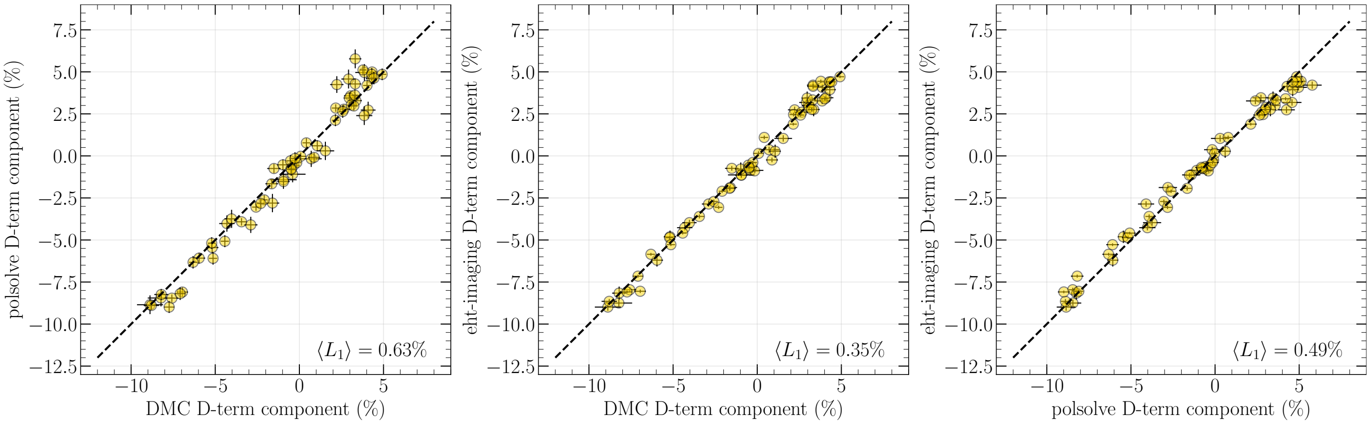

Figure 18. Pairwise comparisons of leakage estimates for ALMA and APEX, obtained from point-source modeling of the intra-site baseline with `polsolve`, `eht-imaging`, and DMC. Both `polsolve` and `eht-imaging` leakages are derived from multi-source fits, while the DMC leakages are derived from fitting to 3C 279 only. Each panel aggregates leakage estimates from both stations (ALMA and APEX), both bands, and all four observing days. Values quoted in the lower right-hand corner of each panel are the uncertainty-weighted mean absolute deviation for the corresponding pair of fits. The dashed line on each plot marks where $y = x$.

LR and LL) at each intermediate frequency band[12]. We can therefore use the high S/N on ALMA–APEX to estimate the D-terms at each intermediate frequency band and study the possible frequency dependence of the instrumental polarization of ALMA and APEX. The results are shown in Figure 19. This test showed very stable D-term estimates across the entire band, motivating band-averaging.

E. FIDUCIAL LEAKAGE D-TERMS FROM M87 IMAGING

We provide fiducial M87 D-term estimates for each method in Table 5. The D-terms for LMT, SMT, and PV are depicted in Figure 2. In Figure 20 we show an example of one-to-one software comparisons of the campaign-average D-terms for LMT, PV and SMT.

F. PRELIMINARY IMAGING RESULTS FOR M87

In this appendix we present the preliminary polarimetric results on M87 obtained using the three imaging methods. These preliminary images were generated "by hand", with manual tuning of free parameters in the imaging and calibration process, before full parameter surveys were done to choose parameters more objectively and evaluate uncertainties. Nonetheless, in this early stage of imaging we found a high degree of similarity in the recovered structure and D-terms between methods; these results guided the design of our synthetic data tests and parameter survey strategy we pursued to obtain our final polarimetric images of M87.

The preliminary polarimetric imaging and leakage calibration used the April 11 pre-processed data set (Stokes $\mathcal{I}$ self-calibrated and leakage correction applied for stations with a co-located partner) are not blind tests in analogy to total intensity imaging (see Paper IV). In this preliminary step, parameters used in the methods were hand-tuned by expert users. This preliminary imaging demonstrates that the total intensity image recovered from the pre-processed data set is roughly consistent with the original Stokes $\mathcal{I}$ image without script fine-tuning.

In Figure 21, we present our recovered total intensity and preliminary polarimetric images of M87 on April 11 produced by the three chosen methods. In Figure 21, we also show D-terms associated with these images. Each method reproduces consistent D-terms for all remaining long-baseline EHT stations. The preliminary polarimetric images are roughly consistent across methods. In all images, the M87 ring-like structure is predominantly polarized mostly in the south–west part with a fractional polarization of about $|m| \sim 15\%$. The EVPAs are organized into a coherent pattern along the ring. However, small differences in fractional polarization and polarized flux density are evident between the three packages. The preliminary results strongly motivate the need for a full parameter survey for each introduced method.

G. DESCRIPTION OF THE PARAMETER SURVEY AND SCORING FOR EACH METHOD

In what follows, we describe each method's approach to surveying the space of free parameters available to it, scoring the results using these six synthetic data models, and from these scores, determining a fiducial set of parameters to use in the final polarimetric imaging of M87. We use each method's fiducial parameter settings to obtain the final M87 images and calibrated D-terms.

[12] In the VLBI correlation, each 2 GHz band is divided into 32 contiguous sub-bands of equal width, which are called "intermediate frequency bands", or IFs .

Table 5. Fiducial set of low-band D-terms for each station as derived from M87 data via polarimetric imaging with fiducial polarimetric survey parameters.

	April 5		**April 6**		**April 10**		**April 11**	
Method	D_R (%)	D_L (%)	D_R (%)	D_L (%)	D_R (%)	D_L (%)	D_R (%)	D_L (%)
				LMT				
`eht-imaging`	$0.48+3.16i$	$-0.47+2.17i$	$1.29+5.37i$	$0.56+2.23i$	$-2.04+2.99i$	$-5.02+0.64i$	$1.43+3.12i$	$-0.44+0.52i$
`polsolve`	$1.5+2.92i$	$-2.26+0.12i$	$1.63+5.1i$	$-0.38+1.48i$	$1.24+0.69i$	$-3.65-2.69i$	$2.48+0.47i$	$-1.12+0.25i$
`LPCAL`	$0.6+2.11i$	$-3.0+0.0i$	$-0.2+4.9i$	$-1.35+0.73i$	$-1.79+2.96i$	$-3.88-2.11i$	$0.74+2.42i$	$-1.37-0.51i$
DMC	$1.8+3.3i$	$-1.5+1.7i$	$2.5+6.4i$	$-0.8+2.4i$	$-2.0+4.3i$	$-5.2+2.3i$	$2.2+4.4i$	$-0.8+0.3i$
	$(0.6+0.6i)$	$(0.6+0.6i)$	$(0.6+0.6i)$	$(0.6+0.6i)$	$(1.1+1.0i)$	$(1.4+1.4i)$	$(0.5+0.5i)$	$(0.5+0.4i)$
THEMIS	$2.3+1.7i$	$-1.0+2.9i$	$3.1+5.5i$	$-0.2+3.0i$	$-0.2+3.5i$	$-4.8+0.8i$	$2.8+2.9i$	$-0.7+0.9i$
	$(0.7+0.8i)$	$(0.7+0.7i)$	$(0.6+0.6i)$	$(0.7+0.8i)$	$(1.4+1.3i)$	$(1.5+1.6i)$	$(0.7+0.7i)$	$(0.8+0.8i)$
				SMT				
`eht-imaging`	$3.94+7.51i$	$-4.84+9.46i$	$3.36+7.78i$	$-4.22+8.19i$	$4.88+9.26i$	$-3.01+9.2i$	$3.84+7.46i$	$-4.38+10.46i$
`polsolve`	$3.99+7.45i$	$-5.42+9.17i$	$3.66+7.62i$	$-4.52+7.49i$	$5.61+8i$	$-5.3+9.68i$	$4.07+7.48i$	$-5.88+10.95i$
`LPCAL`	$3.24+8.23i$	$-5.68+9.04i$	$3.14+8.29i$	$-3.94+7.22i$	$4.51+9.54i$	$-5.6+6.64i$	$4.0+7.56i$	$-5.56+9.3i$
DMC	$2.0+9.4i$	$-3.1+10.7i$	$1.7+10.3i$	$-2.6+8.6i$	$3.1+9.3i$	$-2.9+8.9i$	$3.0+8.8i$	$-4.0+9.5i$
	$(0.9+0.6i)$	$(0.9+0.7i)$	$(0.9+0.6i)$	$(0.9+0.6i)$	$(1.3+1.2i)$	$(1.2+0.9i)$	$(0.8+0.5i)$	$(0.8+0.5i)$
THEMIS	$2.7+8.3i$	$-4.4+9.6i$	$2.8+9.1i$	$-3.1+8.7i$	$1.0+9.3i$	$-3.6+9.7i$	$2.9+6.9i$	$-3.9+10.5i$
	$(0.8+0.8i)$	$(0.8+0.8i)$	$(0.8+0.7i)$	$(0.8+0.7i)$	$(1.3+1.3i)$	$(1.3+1.4i)$	$(0.7+0.7i)$	$(0.7+0.7i)$
				PV				
`eht-imaging`	$-14.57+2.34i$	$15.77+1.21i$	$-13.01+3.51i$	$13.16+1.75i$	$-10.82-2.25i$	$14.68+1.71i$	$-12.70-1.21i$	$14.86+0.57i$
`polsolve`	$-10.64+1.2i$	$13.23+2.83i$	$-13.1+3.99i$	$11.53+2.35i$	$-6.38-1.08i$	$16.64+5.17i$	$-11.32-0.63i$	$13.96+0.84i$
`LPCAL`	$-9.98+0.67i$	$16.4+2.22i$	$-11.66+1.5i$	$14.61+2.14i$	$-12.42-3.85i$	$16.56+3.66i$	$-11.54+0.18i$	$16.16+1.56i$
DMC	$-14.0+1.7i$	$18.2-0.4i$	$-11.9+4.0i$	$12.8-1.1i$	$-11.1-1.0i$	$12.7+3.3i$	$-14.2+0.1i$	$12.9-1.6i$
	$(1.5+1.9i)$	$(1.6+2.2i)$	$(1.3+1.6i)$	$(1.3+1.6i)$	$(2.6+2.5i)$	$(3.1+3.1i)$	$(1.0+1.7i)$	$(1.0+1.6i)$
THEMIS	$-13.9+5.2i$	$17.7-2.4i$	$-10.6+6.3i$	$13.8+1.0i$	$-13.4+0.2i$	$17.1+1.7i$	$-13.6+3.6i$	$14.5-0.3i$
	$(1.5+1.3i)$	$(1.5+1.7i)$	$(2.2+1.5i)$	$(1.3+1.4i)$	$(2.4+1.9i)$	$(2.5+2.4i)$	$(1.1+1.5i)$	$(1.1+1.3i)$
				residual leakage ALMA				
`LPCAL`	$0.91+0.81i$	$0.17+1.82i$	$-0.48+1.18i$	$-0.62-0.58i$	$-2.16+0.94i$	$-1.11-0.74i$	$-0.65+0.83i$	$0.16+0.43i$
DMC	$0.4-0.5i$	$-0.4+3.5i$	$0.1+0.7i$	$0.2+4.1i$	$-0.7+2.5i$	$2.5+1.3i$	$0.3+1.3i$	$2.4+2.5i$
	$(0.7+0.8i)$	$(0.6+0.5i)$	$(0.7+0.7i)$	$(0.5+0.6i)$	$(0.7+0.7i)$	$(1.1+1.0i)$	$(0.6+0.6i)$	$(0.5+0.5i)$
THEMIS	$0.9+1.4i$	$-0.4+3.1i$	$0.7+1.5i$	$0.4+3.6i$	$0.3+0.6i$	$-0.6+2.5i$	$0.1+1.9i$	$1.9+1.9i$
	$(0.7+0.8i)$	$(0.6+0.6i)$	$(0.6+0.7i)$	$(0.6+0.7i)$	$(1.3+1.1i)$	$(1.2+0.9i)$	$(0.6+0.7i)$	$(0.7+0.6i)$
				residual leakage APEX				
`LPCAL`	$1.47+0.22i$	$-0.62-1.69i$	$-0.13-0.40i$	$-1.27-0.56i$	$0.03-1.00i$	$-0.97+1.31i$	$1.12-0.26i$	$-0.45-0.01i$
DMC	$-4.3+1.0i$	$-0.8-0.5i$	$2.8+0.5i$	$-1.1-1.3i$	$7.3-2.6i$	$1.9+2.1i$	$2.4+-0.2i$	$-2.7-0.1i$
	$(2.1+2.1i)$	$(2.0+2.1i)$	$(1.3+1.3i)$	$(1.2+1.2i)$	$(1.7+1.4i)$	$(2.0+2.1i)$	$(1.1+1.1i)$	$(1.0+1.0i)$
THEMIS	$0.6-0.4i$	$-1.7-0.4i$	$3.6+0.2i$	$-1.6-1.8i$	$3.7-1.7i$	$0.5-0.6i$	$2.7+0.8i$	$-1.7-0.6i$
	$(1.2+1.3i)$	$(1.3+1.2i)$	$(0.9+0.8i)$	$(0.8+0.8i)$	$(1.3+1.6i)$	$(1.4+1.6i)$	$(0.8+0.8i)$	$(0.7+0.7i)$
				residual leakage SMA				
`LPCAL`	$-1.59+8.08i$	$7.46+9.51i$	$3.02+8.02i$	$12.24+5.44i$	$9.36+11.73i$	$13.42+14.64i$	$0.21+4.53i$	$14.01+8.40i$
DMC	$0.8-3.4i$	$-4.9-1.0i$	$-1.5-2.1i$	$0.2+0.3i$	$5.5+5.4i$	$-1.8+2.5i$	$-0.2-2.2i$	$0.7+1.9i$
	$(1.1+1.2i)$	$(2.4+2.5i)$	$(1.3+1.4i)$	$(2.7+2.8i)$	$(4.5+5.5i)$	$(3.1+3.3i)$	$(2.5+2.6i)$	$(1.6+1.8i)$
THEMIS	$2.4-2.9i$	$-1.4-0.8i$	$-0.4+0.7i$	$0.3+1.4i$	$2.3+1.0i$	$-4.9+7.1i$	$-2.8+0.4i$	$1.4+3.2i$
	$(1.3+1.2i)$	$(1.2+1.3i)$	$(1.5+1.6i)$	$(1.3+1.7i)$	$(2.8+2.9i)$	$(3.2+3.3i)$	$(1.6+1.7i)$	$(1.9+2.1i)$
				residual leakage JCMT				
DMC	$1.5-0.1i$	...	$1.1-1.8i$	...	...	$-1.3+1.1i$	...	$2.0+0.1i$
	$(1.2+1.2i)$	...	$(1.3+1.4i)$	...	...	$(2.9+3.2i)$	...	$(1.5+1.7i)$
THEMIS	$2.3-0.4i$	...	$0.8+0.3i$	...	...	$-0.6+6.9i$	...	$2.7+1.1i$
	$(1.2+1.3i)$	...	$(1.4+1.5i)$	...	...	$(2.7+2.6i)$	...	$(1.5+1.6i)$

Note – The `eht-imaging`, `polsolve`, and `LPCAL` methods pre-calibrate the ALMA, APEX, SMA, and JCMT D-terms using zero-baseline fitting (see Section 4.2), and so only the D-terms for stations forming long baselines (i.e., LMT, SMT, and PV) are reported for these approaches. DMC and THEMIS do not pre-calibrate the zero-baseline D-terms (see Appendix C.3), and we report here "residual" leakage values – i.e., the excess leakage, as determined by DMC and THEMIS, over that obtained from zero-baseline fitting (given in Table 3 for ALMA and Table 4 for APEX, JCMT, and SMA). The excess leakages from `LPCAL` originates from the fact that `LPCAL` is not able to fix D-terms of specific stations to be certain values. Thus, `LPCAL` obtains solutions for those stations and the non-zero D-terms indicate that there might be possible residual leakages after zero-baseline fitting or uncertainties in `LPCAL`, e.g., from the similarity approximation, are translated into them. The uncertainties for each of the DMC and THEMIS leakage estimates are quoted in parenthesis. D-terms for LMT, SMT, and PV are depicted in Figure 2.

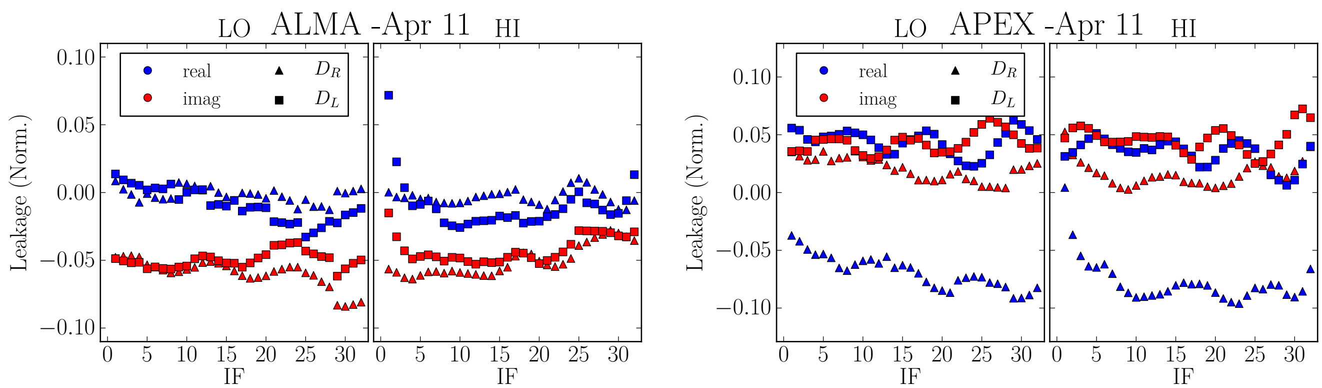

Figure 19. ALMA (left) and APEX (right) D-term spectra recovered on April 11. Each band has a width of 2 GHz and is divided into 32 intermediate frequency sub-bands (IFs) of equal width.

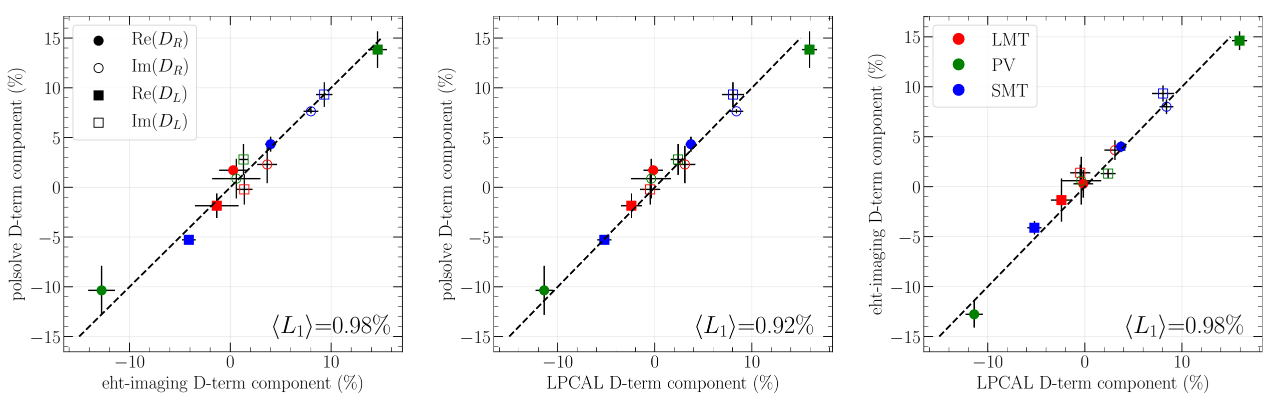

Figure 20. Example one-to-one software comparisons of the campaign-average D-term estimates for LMT, PV, and SMT polarimetric imaging and leakage calibration of M87 observations. Left: Comparison of `eht-imaging` estimates against `polsolve` estimates. Center: Comparison of `LPCAL` estimates against `polsolve` estimates. Right: Comparison of `LPCAL` estimates against `eht-imaging` estimates. Norm L_1 is averaged over left, right, real, imaginary components of the D-terms and over all shown EHT stations. See section 4.2 for averaged Norm L_1 between `eht-imaging`/`polsolve`/`LPCAL` and THEMIS/DMC.

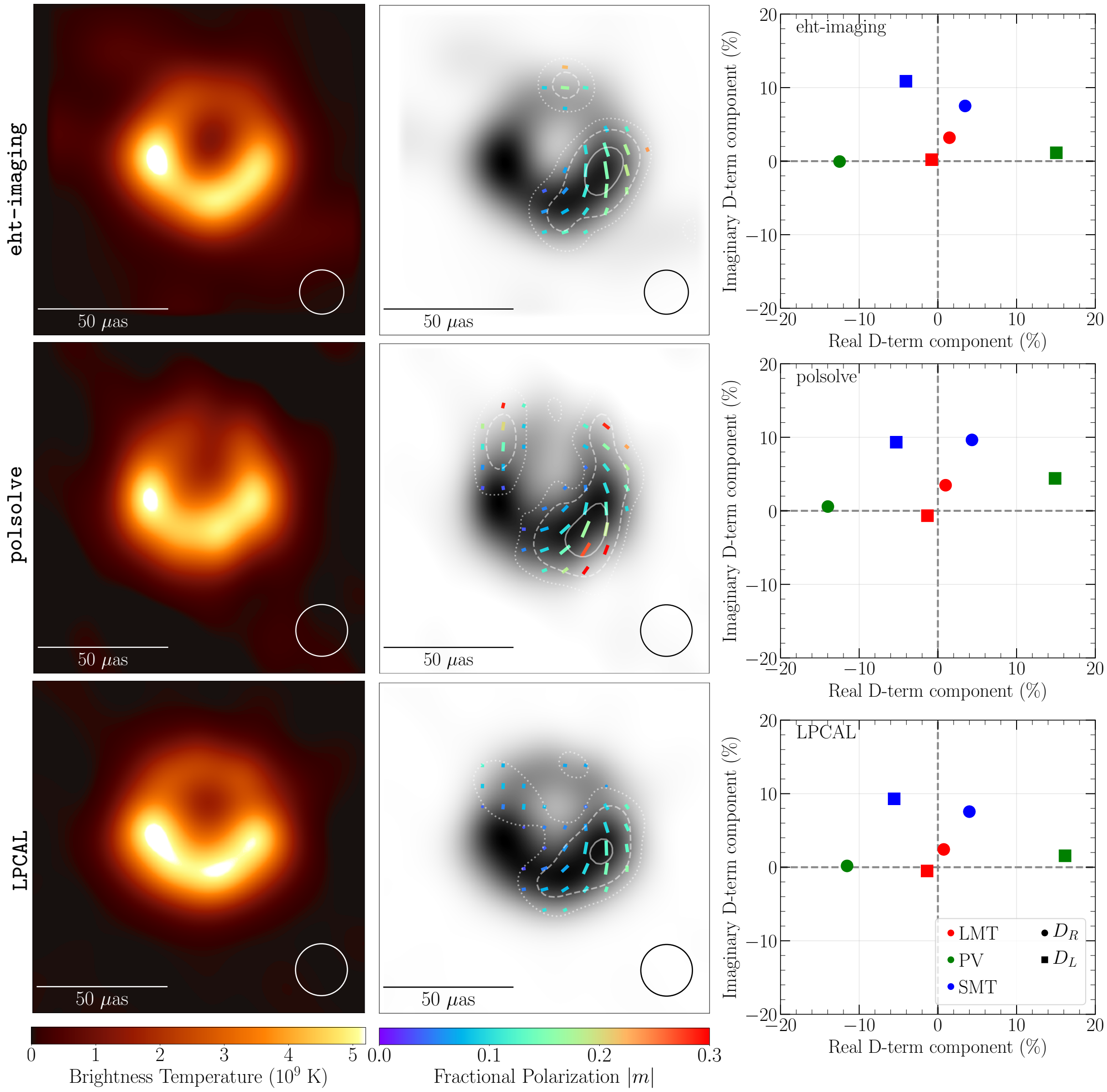

Figure 21. *Left:* Preliminary April 11 total intensity images reconstructed with `eht-imaging`, `polsolve`, and `LPCAL`. `eht-imaging` images are blurred with a 17.1 μas circular Gaussian, to obtain an equivalent resolution to the `polsolve` and `LPCAL` CLEAN images restored with a 20 μas circular Gaussian. *Middle:* Corresponding polarimetric reconstructions obtained as a result of the full-array leakage calibration. Total intensity is shown in the background in grayscale. Polarization ticks indicate the EVPA, the tick length is proportional to the linear polarization intensity magnitude, and color indicates fractional linear polarization. The contours mark linear polarized intensity. The solid, dashed, and dotted contour levels correspond to linearly polarized intensity of 20, 10, and 5 μJy/μas^2. Cuts were made to omit all regions in the images where Stokes $\mathcal{I} < 10\%$ of the peak flux density and $p < 20\%$ of the peak polarized flux density. In all reconstructions, the region with the highest linear polarization fraction and polarized intensity is predominantly in the south–west portion of the ring. *Right*: Preliminary D-terms for SMT, PV, and LMT derived via leakage calibration through `eht-imaging`, `polsolve`, and `LPCAL` polarimetric imaging.

G.1. *eht-imaging parameter survey*

The polarimetric imaging procedure alternates between imaging via minimization of the objective function (Equation C6) and D-term calibration, as described in Section C.2. In the imaging stage, the critical parameters that influence the final reconstruction include the four hyperparameters $\alpha_P, \alpha_m, \beta_{HW}$, and β_{TV} that set the relative weighting in the objective function between the different data constraints and regularizing terms. In surveying different parameters in the `eht-imaging` survey, we fix $\alpha_m = 1$ and vary the other three hyperparameter weights.

In addition to the hyperparameter weights, an additional free parameter in our objective function is the amount of additional systematic noise to add to the data as a budget for non-leakage sources of systematic errors

Table 6. The parameters surveyed by `eht-imaging`.

Parameter	Values				
α_P	0.01	0.1	**1**	10	100
β_{HW}	0.01	0.1	1	10	**100**
β_{TV}	0.01	0.1	**1**	10	100
$n_{\rm iter}$	1	10	**50**	...	...
$f_{\rm sys}$	0	0.002	**0.005**	0.01	...

Note – The selected fiducial parameters are displayed in bold.

(see Paper III, Paper IV). To account for these systematics, we add a term equal to $f_{\rm sys} \times |\tilde{\mathcal{I}}|$ in quadrature to our baseline thermal noise estimates, where $f_{\rm sys}$ is an overall multiplicative factor. Finally, we also include as a parameter in our surveys the number of iterations $n_{\rm iter}$ of alternating between imaging and calibrating the station D-terms. The full list of parameters we vary in the `eht-imaging` parameter survey is listed in Table 6, with the fiducial parameters used in reconstructing images of M87 denoted in bold.

To select a fiducial set of parameters that performs best on all six synthetic data tests, we assign each parameter set $\mathbf{p}$ two scores on each synthetic data set a; one scoring the fidelity of the polarized image reconstruction $s_{\rm fid}(\mathbf{p})$, and one scoring the accuracy of the D-term estimation $s_{\rm dterm}(\mathbf{p})$. First, we score the final image fidelity by computing the normalized cross correlation $\rho_{\rm NX}$ between the reconstructed and ground truth polarimetric intensity distribution. That is, we use equation 15 of Paper IV on the images of $\sqrt{\mathcal{Q}^2 + \mathcal{U}^2}$. Then the fidelity score for the parameter set $\mathbf{p}$ on the synthetic dataset a is simply

$$s_{{\rm fid},a}(\mathbf{p}) = \rho_{\rm NX}. \tag{G9}$$

We compute the D-term estimation accuracy metric by first calculating the average ℓ_2 distance d_D between the reconstructed D-terms and the ground truth for the data set. For M87, we average this distance over the three stations we calibrate at this stage: SMT, LMT, and PV. Then we transform this distance to a score between 0 (bad) and 1 (good) by using a sigmoid function;

$$s_{{\rm dterm},a}(\mathbf{p}) = 1 - {\rm Erf}\left[d_D/\sqrt{2}d_{\rm tol}\right], \tag{G10}$$

where $d_{\rm tol}$ is a threshold for the average distance between the ground truth and recovered D-terms beyond which we begin to heavily penalize the reconstruction. We set $d_{\rm tol} = 5\%$

Finally, having computed the two scores $s_{{\rm fid},a}$ and $s_{{\rm dterm},a}$ for the parameter set $\mathbf{p}$ on the synthetic data set a, we compute a final score $s(\mathbf{p})$ for the parameter set by multiplying these individual scores together on all six synthetic data sets a:

$$s(\mathbf{p}) = \prod_a s_{{\rm fid},a} s_{{\rm dterm},a}. \tag{G11}$$

We then have a final score s for each parameter set incorporating its performance in accurately reconstructing the polarized flux distribution and input D-terms on six synthetic data sets. We take the parameter set $\mathbf{p}$ with the highest score as our fiducial parameter set.

G.2. *`polsolve` parameter survey*

The `polsolve` algorithm is completely characterized with a few degrees of freedom: the pixel's angular size (smaller values increase the astrometric accuracy of the CLEAN components and, hence, the quality of the deconvolution), the field of view (the effect of this parameter is minimum if a CLEAN masking is applied), the visibility weighting (mainly defined with the Briggs "robustness" parameter, r; Briggs 1995), and the division of the Stokes $\mathcal{I}$ model into sub-components of constant fractional polarization (see Section C.1).

The first step in the `polsolve` procedure is to generate a first version of the $\mathcal{I}$ image (using the CASA task `clean`). Several iterations of phase and amplitude self-calibration (using tasks `gaincal` and `applycal`) may be applied to the data, in order to optimize the dynamic range of the $\mathcal{I}$ model. The self-calibration gains are forced to be equal for the R and L polarizations at all antennas[13]. Then, the $\mathcal{I}$ model is split into several sub-components and formatted for its use in `polsolve`, using the CASA task `CCextract`[14]. Finally, `polsolve` estimates the D-terms of LMT, SMT and PV, together with the fractional polarizations and EVPAs of all the source sub-components. The estimated D-terms are applied to the data and a final run of `clean` is peformed, to generate the final version of the full-polarization images.

In the `polsolve` parameter survey, the $\mathcal{I}$ division is done in two ways. On the one hand, a centered square mask of $50{\times}50\,\mu$as is created and divided into a regular grid of $n \times n$ cells (in this case, cells that do not contain CLEAN components are not used in the fit). On the other hand, a centered circular mask of $40{-}80\,\mu$as diameter is created and divided azimuthally into a regular set of n pieces. The full parameter survey with `polsolve` consists of an exploration of:

1. Both types of model sub-divisions (i.e., either regular grid or azimuthal cuts), with n running from 1 to 12.

2. Robustness parameter, from $r = -2$ to $r = 2$ (i.e., from uniform to natural weighting) in steps of 0.2.

[13] A necessary assumption at this stage, since polarization-dependent gains do not commute with the Dterm corrections already applied to ALMA, APEX and SMA

[14] Part of the `CASA-poltools` software package developed by I. Martí-Vidal: `https://code.launchpad.net/casa-poltools`

3. Relative weight of the ALMA antenna (which affects the shape of the point spread function, PSF, of the instrument, especially for values of r far from -2), running from 0.1 to 1.0 in steps of 0.1.

4. In all images, a circular CLEAN mask of diameter varying from 40 to 80 μas (in steps of 5 μas) is used. The size of the CLEAN mask is the same as the mask to define the $\mathcal{I}$ model sub-division.

The pixel size is fixed to 1 μas and the image size covers 256×256 μas. We note that for models with extended emission (i.e., Models 1 and 2, see Figure 4) an additional CLEAN mask is added at the southern part of the image.

For each combination of parameters in the survey, we compute the L_1 norm of the differences between true and estimated D-terms, as well as the correlation coefficients, ρ_s (for each Stokes parameter, s) between the CLEAN image reconstructions and true source structures, properly convolved with the same beam. These quantities can be used to select the best combination of parameters for an optimum D-term calibration and image reconstruction (the fiducial imaging parameters for `polsolve`). Depending on the relative weight that is given to L_1 and the average image correlation, $\rho = (\rho_\mathcal{I} + \rho_\mathcal{Q} + \rho_\mathcal{U})/3$, we obtain slightly different fiducial parameters.

In Figure 22, we show an example plot from our `polsolve` parameter survey for Model 5 (see Figure 4, similar results are found for the rest of models in the survey). The chosen figure of merit is $(1-\rho)L_1$ (where ρ and L_1 are related to the image fidelity and Dterm deviation, respectively; see Sect. 5.1), which we show as a function of the robustness parameter and the number of slices in the azimuthal model sub-division.

From Figure 22, we note that the dependence of the figure of merit with n becomes weak for values of n larger than $3-4$ and robustness parameters between -1 and -2. This behavior also happens if the regular gridding is used to generate the $\mathcal{I}_i$ sub-components. A qualitative explanation of this effect may be that large values of n translate into sub-components of sizes smaller than the synthesized resolution. Therefore, increasing the number of sub-components does not improve the fit, since the small separations between neighboring sub-components correspond to spatial frequencies that are not sampled by the interferometer. Conversely, the fitted fractional polarizations of close by sub-components become highly correlated in the fit, and the L_1 norm of the D-terms saturate around a minimum value.

Based on the combined analysis of all the synthetic datasets, we determine the fiducial parameters for `polsolve`: a robustness parameter of -1 (though -2 produces similar results, especially for models 4 to 6, see Figure 4), a circular slicing with $n = 8$ subcomponents (which produces results similar to $9-10$ sub-components, and also similar to those from a regular gridding, $n \times n$, with $n = 3-5$), relative ALMA weights of 1.0 (which produces similar results also for values between $0.5-1.0$), and a circular CLEAN mask of 50 μas diameter. The range of values for L_1 and the correlation coefficients among images (obtained from these fiducial parameters) are shown in Figs. 4 and 5.

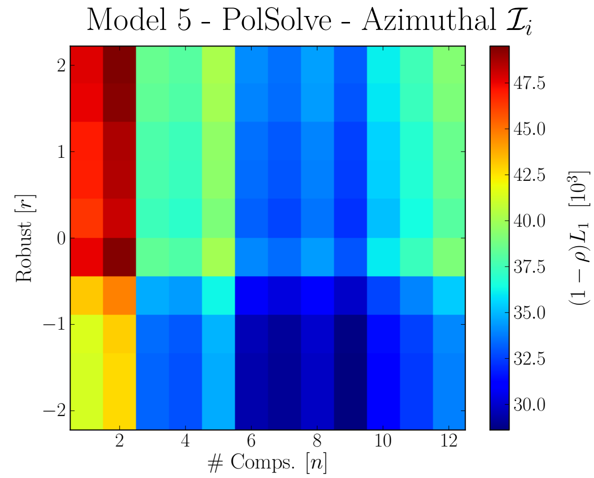

Figure 22. Figure of merit $(1-\rho)L_1$ (lower values mean better results) for the `polsolve` algorithm (running on Model 5), as a function of its two main degrees of freedom (i.e., $\mathcal{I}$ sub-division and visibility weighting).

G.3. *LPCAL parameter survey*

The standard procedure for D-term calibration using `LPCAL` is as follows. 1. Select a calibrator which has either a low fractional polarization or a simple polarization structure, and a wide range of parallactic angle coverage. This is M87 in our case. 2. Produce a total intensity CLEAN map of the calibrator with e.g., `Difmap`. 3. Split the CLEAN model into several submodels with the task CCEDT in AIPS. `LPCAL` assumes that each sub-model has a constant fractional polarization and EVPA. Therefore, the more sub-models we use to divide the Stokes I image, the more degrees of freedom we have for modelling the source linear polarization. 4. Run `LPCAL` using the sub-models.

We follow this standard procedure for the D-term calibration using M87. We consider an additional parameter, the ALMA weight scaling factor, for this work. Down-weighting the ALMA can be useful when there are significant systematic uncertainties in the ALMA visibilities. In this case, the solutions for other stations can be distorted as the fitting would be dominated by the ALMA baselines due to its high sensitivity and the corresponding smaller formal error bars. In addition to the ALMA weight scaling factor, we consider the number of CLEAN sub-models as a main parameter that may affect D-term estimation with `LPCAL` significantly.

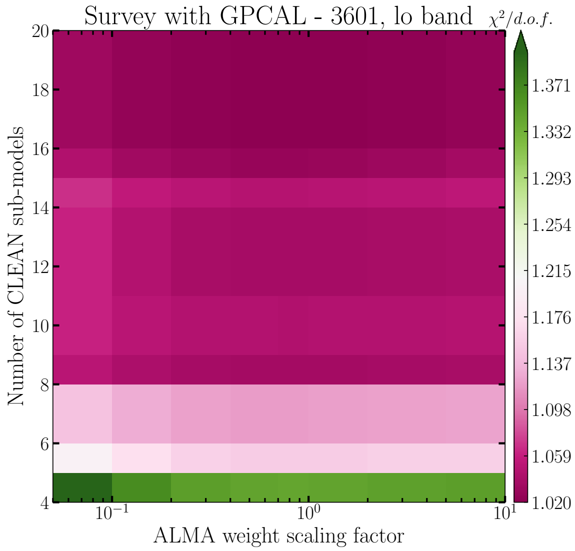

Figure 23. Distribution of the reduced chi-squares of the fits with `GPCAL` depending on the ALMA weight scaling factor and the number of CLEAN sub-models. `GPCAL` parameters which **do** produce nearly identical results to `LPCAL` were used. The M87 real data for April 11 and **low band** was used for the survey. The larger number of sub-models provides a better fit, while the results are insensitive to the ALMA weight scaling factors.

We first performed a manual parameter survey using the synthetic data. We reconstructed D-terms with `LPCAL` by using different numbers of sub-models and ALMA weight scaling factors, and compared with the ground truth values. We conclude that a relatively large number of sub-models ($\gtrsim$ 10) gives better reconstructions, while the results do not change much when more than ten sub-models are used. Also, we find that the results are not sensitive to the ALMA weight re-scaling. The strategy and parameters we adopted could reproduce the ground-truth D-terms in the synthetic data within an accuracy of $\sim 1\%$ (Figure 5).

We further investigated the effects of these parameters on the real data. We employed `GPCAL`, by using the parameters which can produce nearly identical results with `LPCAL` and provide the reduced chi-square ($\chi^2_{\rm red}$) of the fits, to examine the improvements of the statistics with changing the parameters. This is based on the fact that we found a correlation between the goodness of the D-term reconstructions and $\chi^2_{\rm red}$ (i.e., χ^2 divided by the number of degrees of freedom) using the synthetic data with `GPCAL`. Figure 23 shows the distribution of $\chi^2_{\rm red}$ for the two parameters from the survey on the M87 data for day 3601. We explored the number of sub-models from 5 to 20 with an increment of one and the ALMA weight scaling factor of (0.1, 0.2, 0.4, 0.7, 1.0, 2.0, 5.0, 10.0). Similar to our conclusions from the synthetic data analysis, we found that the statistics improve with a larger number of sub-models up to about ten sub-models. Also, the result is insensitive to changing the ALMA weight scaling factors. This trend was seen for the M87 data for the other days as well. Therefore, we conclude that the parameters we used for the real data analysis with `LPCAL` are reasonable.

Based on these results, we analyzed the real M87 data. The Stokes $\mathcal{I}$ image is reconstructed with CLEAN in `Difmap`. The CLEAN components are divided into several sub-models by the task CCEDT in AIPS. The D-terms for all antennas are solved by using `LPCAL` in AIPS. This includes the possible residual D-terms of ALMA, APEX, and SMA. This procedure was applied independently by many co-authors to test its robustness. We let each person use different schemes for CLEAN and different parameters. For example, different (u, v) weighting parameters, CLEAN windows, and CLEAN cutoffs were used. Some people downweighted ALMA baselines through ParselTongue (Kettenis et al. 2006), a Python interface to AIPS, while others did not. Also, some people averaged the data in time for `LPCAL`, while others did not. Then, we obtained the representative D-terms by taking the median of the real and imaginary parts of the D-terms of each station, respectively. This approach allows us to take into account the uncertainties in `LPCAL` that may be associated with the Stokes $\mathcal{I}$ image reconstruction and the parameters used for the different tasks. One exception is that all people split their CLEAN models into a number of sub-models ($\gtrsim$ 10), in accordance with our parameter survey results.

G.4. *DMC parameter survey*

For the DMC image reconstructions, we surveyed two hyperparameters: the pixel separation and the image field of view. Because the DMC method fits for a systematic uncertainty term in addition to the image and calibration parameters, all fits are formally "good" from the perspective of, e.g., a χ^2 metric. We thus determine an acceptable fit for a particular dataset to be the one that minimizes the number of model parameters (i.e., a combination of largest pixel separation and smallest image field of view) while recovering the expected level of systematic uncertainty (i.e., 0% for the synthetic datasets and 2% for the M87 datasets; see Paper III) within some threshold (taken to be the 3σ bounds determined by the posterior distribution).

G.5. Themis *parameter survey*

Associated with Themis reconstructions are two hyperparameters corresponding to the two-dimensional number of control points. These have natural values set by the number of independent modes that may be reconstructed; for the EHT, 5×5. Because this resulted in formally acceptable fits, i.e., reduced-χ^2 near unity, and based on similar considerations from Stokes I image reconstructions (see, e.g., Broderick et al. 2020a), no additional exploration was required for M87. For the

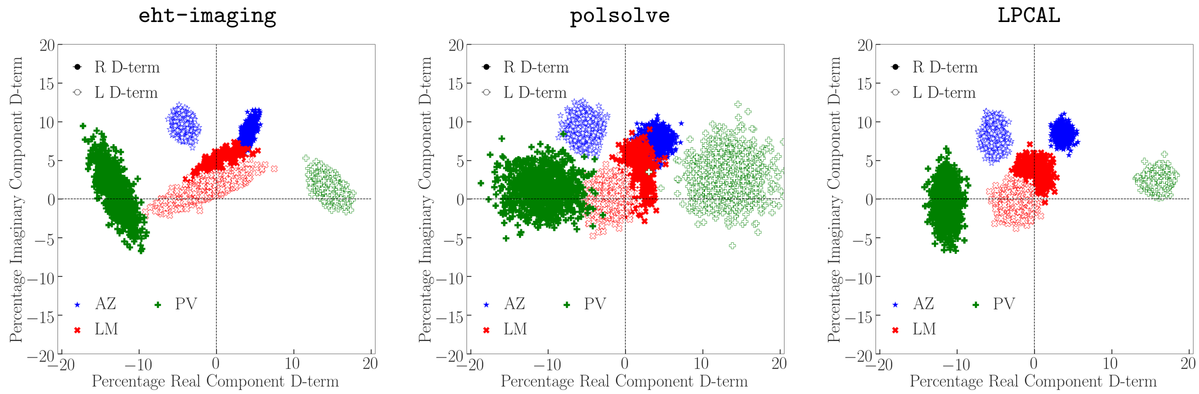

Figure 24. Samples of 1000 SMT, LMT, and PV D-terms applied to the data and used in each of the three image reconstruction algorithms in the Monte-Carlo procedure described in Section 5.2. Each D-term was drawn from a normal distribution with no correlations between the D-terms. The means and (co)variances of these distributions for each method were determined from the set of four fiducial D-term solutions found across the four observing days for each method. `eht-imaging` included covariance between the real and imaginary parts of each D-term in its approach, while `LPCAL` and `polsolve` did not.

Model 1 and 2 synthetic data reconstructions, the number of control points were incrementally increased until acceptable fits were obtained.

H. D-TERM MONTE CARLO SURVEY DETAILS

In this Appendix, we discuss the procedure we follow to assess the consistency of the images to uncertainties in the D-term calibration from the three non-MCMC image reconstruction methods (`eht-imaging`, `LPCAL`, and `polsolve`). We use this method to generate a sample of 1000 images with different D-term calibration solutions for each method on each observing day. These samples are used assess our uncertainty in the polarimetric image structure in Section 5.2 and Section 5.3.

H.1. *Method*

Our method is to use a simple Monte-Carlo approach, similar to the analysis of Martí-Vidal et al. (2012) to study the coupling between antenna D-terms and the recovered polarimetric source structure. For each method, we draw 1000 random sets of D-terms from normal distributions with means and (co)variances determined by considering the fiducial results from Table 5 across the four observing days. We assume for this test that the uncertainties in the D-terms are uncorrelated from station to station and between LCP and RCP. This represents a worst-case scenario test, since correlations between the D-terms would reduce the volume of the D-term parameter space surveyed by each method. The full sample of 1000 D-term sets sampled for the SMT, LMT, and PV on April 11 for `eht-imaging`, `polsolve`, and `LPCAL` are shown in Figure 24.

After drawing a given set of random D-terms, we then applied this calibration solution to the data and reconstructed a polarized image using the same procedure and fiducial imaging parameters selected from the parameter survey for each method in Appendix G. Our imaging scripts in this stage differ from those considered in Appendix G, however, because they do not involve any leakage calibration but only reconstruct the Stokes $\mathcal{Q}$ and $\mathcal{U}$ from the visibilities with the assumed calibration solution applied. That is, we draw a set of random D-terms from distributions reflecting our uncertainty in the recovered D-terms from the earlier parameter survey stage, and then reconstruct an image assuming this D-term calibration is perfect with no further leakage calibration. In contrast, the station amplitude and phase gains are calibrated iteratively in the Stokes $\mathcal{I}$ imaging stage as before.

H.2. *Distributions of image-averaged parameters*

In Figures 25 and 26 we show histograms over each imaging method's sample of 1000 images with different D-term calibration solutions on all four observing days of the four image-integrated quantities used in subsection 5.3. Figure 25 shows histograms of the image net polarization fraction $|m|_{\rm net}$ (Equation 12, plotted in red) and the intensity-weighted average polarization fraction at 20μas resolution $\langle m \rangle$ (Equation 13, plotted in green). Figure 26 shows the amplitude $|\beta_2|$ (plotted in brown) and phase $\angle\beta_2$ (plotted in purple) of the β_2 coefficient of the azimuthal decomposition defined in Equation 14. Because the observations on April 5 and 11 have the highest-quality (u, v) coverage and bracket the observed time evolution of the source, we choose to define acceptable ranges for these parameters (the shaded bars in Figures 25, 26, taken from Table 2) using only these two days. In particular, the poor quality of the April 10 (u, v) coverage leads to broader distributions of the four key quantities with large systematic uncertainties between imaging methods (third columns of Figures 25, 26). The distributions on April 5 and 11 are summarized with mean and $1-\sigma$ error bars in the main text Figure 9, and discussed in Section 5.3.

Finally, Table 7 presents ranges of the image-integrated Stokes parameters $\mathcal{I}, Q, U$ derived from the

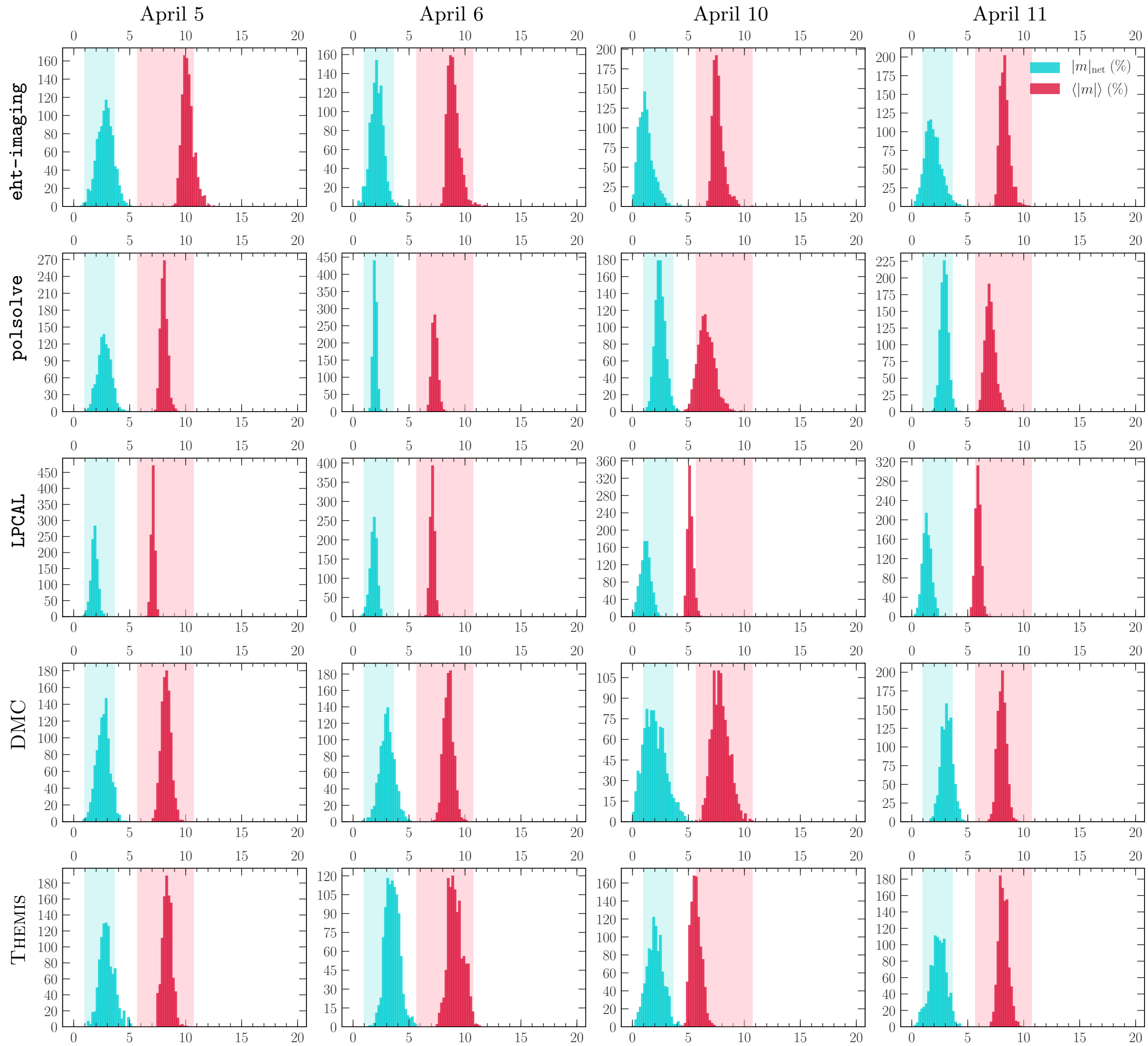

Figure 25. Histograms of the net polarization fraction $|m|_{\rm net}$ (green: Equation 12) and the image-averaged polarization fraction $\langle|m|\rangle$ (red: Equation 13) from each method's survey over random D-term calibration solutions. From left to right, the four columns show histograms for April 5, 6, 10 and 11. In all panels the shaded bands represent the final parameter ranges reported in this work, incorporating the uncertainty both across D-term realizations and reconstruction methods. These ranges are presented in Table 2. Note that as a consequence of the poor (u, v) coverage and parallactic angle sampling, the April 10 image reconstructions from all methods show more systematic uncertainty in the derived parameters than on the other days.

surveys over different D-term calibration solutions from each day of observations. The ranges in Table 7 were calculated by taking the minimum mean -1σ and maximum mean $+1\sigma$ point from the five individual method surveys on each day.

I. CONSISTENCY OF LOW AND HIGH BAND RESULTS FOR M87

The results shown in the main text were obtained for the EHT low band data (centered at 227.1 GHz). In this Appendix, we verify the consistency of these results with the EHT high band data (centered at 229.1 GHz) by repeating several steps of the low band analysis with the same methodology.

I.1. *Fiducial High Band Images and D-terms*

We first compare fiducial D-terms and fiducial polarimetric images derived from the high band data with the low band results reported in the main text (Section 5.1). To produce the high band images, each imaging method

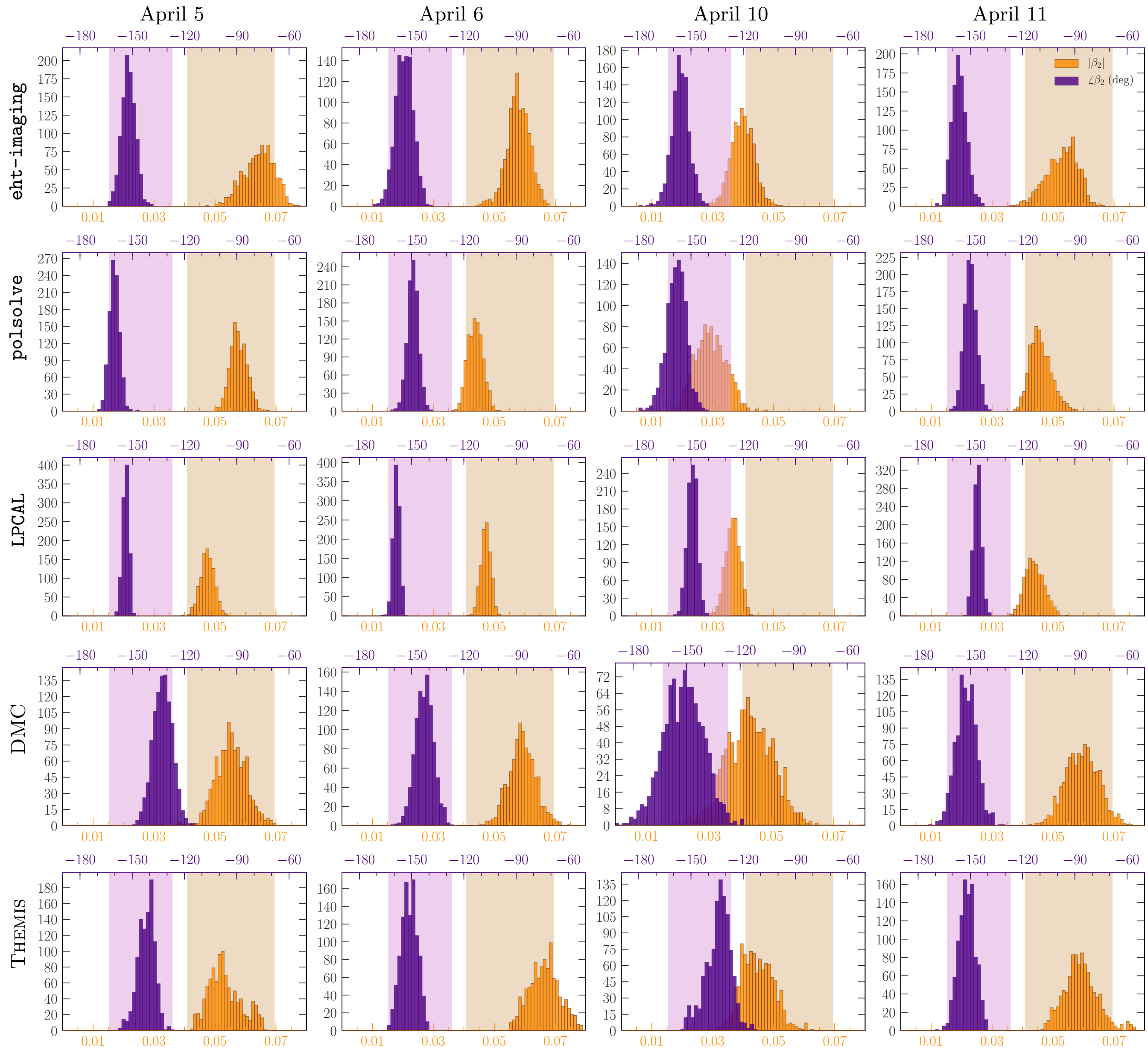

Figure 26. Histograms of the amplitude $|\beta_2|$ (brown: bottom axis) and phase (purple: top axis) $\angle\beta_2$ of the $m = 2$ azimuthal mode of the complex polarization brightness distribution (Equation 14) from each method's survey over random D-term calibration solutions. From left to right, the four columns show histograms for April 5, 6, 10 and 11. In all panels the shaded bands represent the final parameter ranges presented in Table 2. Note that as a consequence of the poor (u, v) coverage, the April 10 image reconstructions from all methods show more systematic uncertainty in the derived parameters than on the other days.

used an identical imaging procedure as for the low band images shown in Figures 6-7 – we did not re-derive parameters specific to each imaging method or repeat the synthetic data surveys described in Section 4.3. This ensures that the methods, while not tuned to high band data, are able to reproduce our most robust results.

Figure 27 shows the low and high band D-terms for LMT, PV, and SMT derived for all five methods. The high band D-terms sit within the systematic scatter among methods in the low band results. Figure 28 compares the method-averaged high and low band consensus images on all four days. The overall level of linear polarization and azimuthal polarization pattern is consistent between the bands on each day.

I.2. *Image-averaged quantities*

To evaluate the consistency of the image-averaged quantities, we extend the analysis presented in Section 5.3 to the high band data. In particular, we generate a sample of 1000 images from the high band data

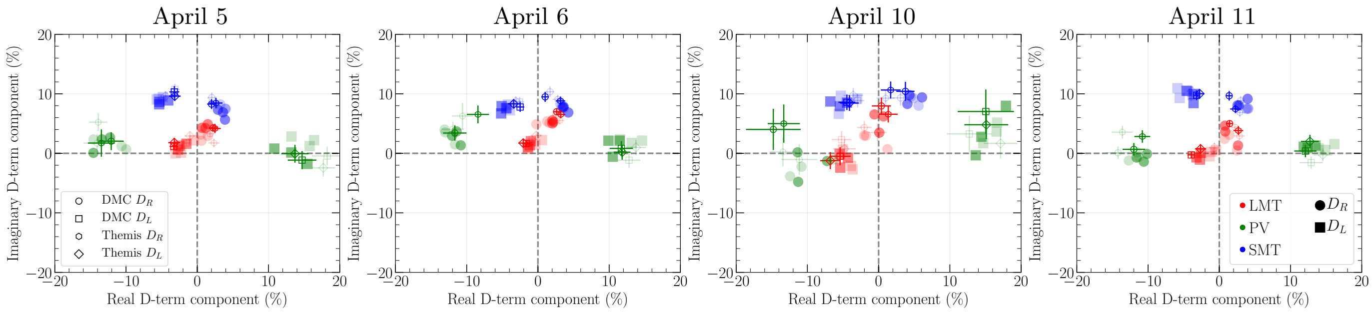

Figure 27. Fiducial D-terms for LMT, PV, and SMT derived via leakage calibration through `eht-imaging`, `polsolve`, `LPCAL` polarimetric imaging and DMC/Themis posterior exploration of M87 data. The D-terms derived from low band (lighter points) and high band (heavier points) are consistent with one another within the systematic scatter among methods seen in the low band results. All D-terms are displayed in the same manner as in the right panels in Figure 2.

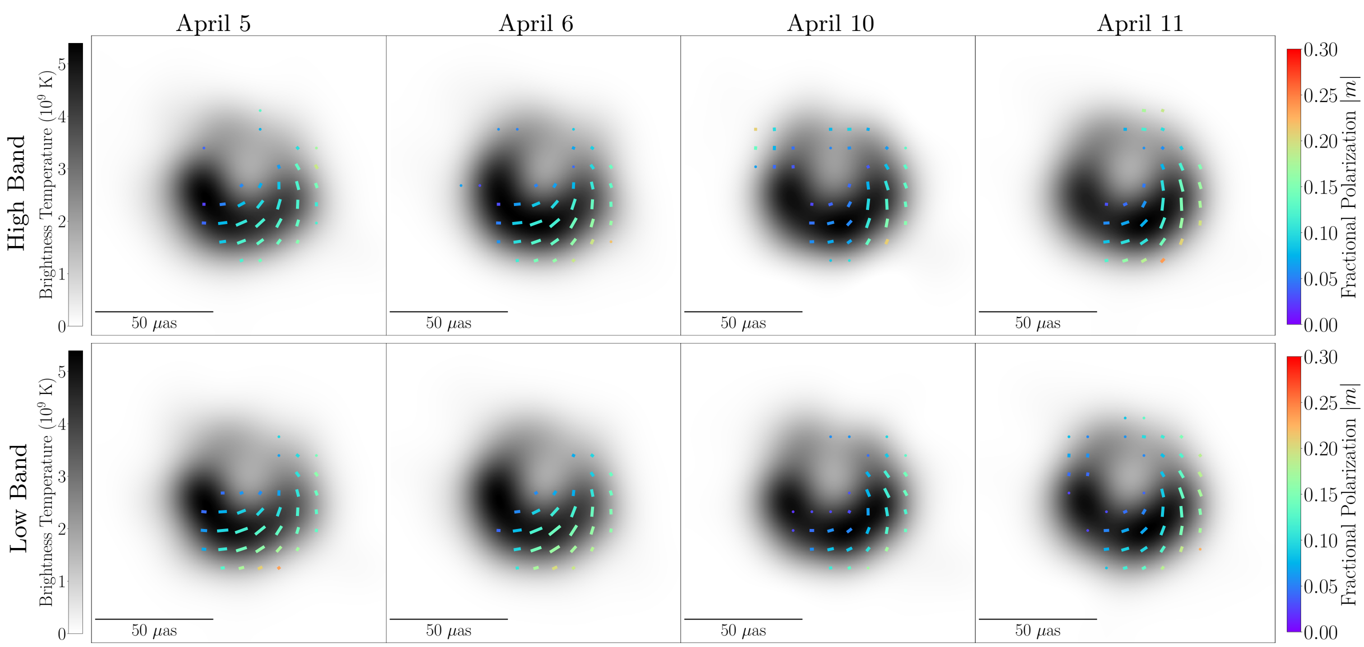

Figure 28. Fiducial M87 average images per day produced by averaging results from our five methods (see Figure 6). Method-average images for all four M87 observation days are shown, from left to right. The top and bottom rows show high and low band results, respectively. The images are all displayed with a field of view of 120 μas, and all images were brought to the same nominal resolution by convolution with the circular Gaussian kernel that maximized the cross-correlation of the blurred Stokes $\mathcal{I}$ image with the consensus Stokes $\mathcal{I}$ image of Paper IV. Total intensity, polarization fraction, and EVPA are plotted in the same manner as in Figure 6.

for each method, exploring a range of different D-term calibration solutions (e.g., Figure 24).

Figure 29 compares results for the key image-integrated metrics ($|m|_{\rm net}$, $\langle m \rangle$, β_2; see definitions in Section 5.3) derived from such image samples. High band results for a given method are generally consistent within 1-2σ with their low-band counterparts, with the largest offset seen on April 11 for the `polsolve` measurements of $|m|_{\rm net}$, $|\beta_2|$ and $\angle\beta_2$. The median high band results all fall within the ranges established using the low band images (Table 2), but the high band median-1σ points fall outside the established ranges for the `eht-imaging` and `polsolve` $|m|_{\rm net}$ measurements on April 11 and for the `polsolve` measurement of $|\beta_2|$ on April 5. Because the imaging procedures and results for low band were extensively validated with synthetic data tests, we choose to use the low band results only in defining the parameter ranges in Table 2. Note that $|m|_{\rm net}$ in particular can be quite sensitive to the choice of imaging hyperparameters used, and these parameters were not re-derived for the high band data.

Figure 30 compares the image-integrated EVPA measured for high and low band across all four days. For April 6, 10, and 11, the results are consistent for each

Table 7. Ranges of image-integrated Stokes $\mathcal{I}, \mathcal{Q}, \mathcal{U}$ obtained by combining each method's D-term calibration survey.

Stokes	Min	Max	Min	Max
	April 5		**April 6**	
$\mathcal{I}$ (Jy)	0.419	0.512	0.376	0.508
$\mathcal{Q}$ (Jy)	−0.0136	−0.0002	−0.0056	0.0075
$\mathcal{U}$ (Jy)	−0.0157	−0.0055	−0.0167	−0.0063
	April 10		**April 11**	
$\mathcal{I}$ (Jy)	0.381	0.545	0.407	0.565
$\mathcal{Q}$ (Jy)	−0.0067	0.0158	0.0030	0.0180
$\mathcal{U}$ (Jy)	−0.0074	0.0115	−0.0113	0.0007

method within the 1σ errorbars derived from the 1000 image sample. As discussed in the main text for the low band results (Figure 30), the EHT images have net EVPAs that are offset from the ALMA-only measurements on arcsecond scales by 30-60 degrees.

On April 5, we see signatures of a systematic EVPA offset between the low and high band images of M87. All five methods show a shift $\Delta\chi_{|\mathrm{HI-LO}|} \approx 20 - 30\,\mathrm{deg}$. However, for each method, the magnitude of the EVPA discrepancy is $\lesssim 2\sigma$. When interpreting this systematic EVPA offset as Faraday rotation, it would correspond to $|RM| \sim 1 - 2 \times 10^7 \mathrm{rad/m}^2$. One can adopt this number as a rough upper limit on resolved rotation measure.

While on the other three days (April 6, 10, and 11) there is no signature of an offset in the image-integrated EVPA, we do see intriguing offsets in the EVPAs in some portions of the resolved images. Such non-uniform rotations may be indicative of Faraday rotation occurring internally in the compact source, but because of the low significance of these detections we make no further effort to interpret them here. Future EHT observations with larger bandwidths and more complete (u, v) coverage will offer an opportunity to investigate offsets in the resolved EVPA structure between the frequency bands in more detail and with increased confidence.

J. LMT, SMT AND PV D-TERMS USING CALIBRATOR DATA: SYNTHETIC DATA TESTS, EXPECTED UNCERTAINTIES AND CONVERGENCE WITH M87 RESULTS

Together with M87, full-array polarimetric calibration and imaging was also attempted on three other sources: 3C 279, observed contemporaneously with M87; J1924–2914 and NRAO 530, observed contemporaneously with our second EHT primary target, Sgr A*, in the second half of each observing day. 3C 279 was observed on the same four days as M87, with the latter two days having the best (u, v) coverage with the addition of SPT. J1924–2914 was observed on all five days of the EHT campaign (the same four days as M87 with the addition of April 7), and NRAO 530 was observed on the first three days of the campaign (April 5–7). Coverage and data quality vary from day to day, depending on the structures of the observations and, in the case of the Sgr A* calibrators, whether ALMA is observing. For optimal calibration and imaging, we make an initial cut based on (u, v) and field rotation angle coverage, and the presence of ALMA in the array. We exclude J1924–2914 and NRAO 530 observations on April 5, which do not have ALMA, and the April 10 two-scan snapshot observations of J1924–2914, which severely lack coverage.

In Figure 31, bottom row, we show the field angle coverage on the three calibrators for their best-coverage day (April 11 for 3C 279 and J1924–2914, and April 7 for NRAO 530). The field angle coverage for M87 on April 11 is shown as well for comparison. Compared to M87, the three calibrators are at sufficiently low declination to also be observed by the SPT, but the elevation stays constant for sources viewed from the South Pole and only a constant field angle is sampled. In Figure 31, top row, we present the $|\breve{m}|$ structure in the (u, v) plane prior to D-term calibration for the best-coverage days of the calibrators; April 11 M87 is also shown for reference. High polarization fraction is expected in M87 on baselines that probe our visibility minima in total intensity, but the source overall is weakly polarized. 3C 279, on the other hand has multiple baselines exhibiting high polarization fraction. The recovery of D-terms for a highly polarized source like 3C 279 would require an extremely accurate source model in both total intensity and polarization. However, 3C 279's complex structure in both total and polarized intensity add to the difficulty of imaging and calibrating the source (Kim et al. 2020). Furthermore, it was found that 3C 279 may have non-negligible Stokes $\mathcal{V}$ as measured by interferometric-ALMA measurements taken contemporaneously to our EHT campaign (see Appendix E.2 in Goddi et al in prep 2020), which breaks our Stokes $V = 0$ assumptions in our calibration and imaging software. Based on these findings, 3C 279 is thus not the best choice for D-term comparisons with M87.

J1924–2914 and NRAO 530 exhibit low polarization fractions on most baselines (Figure 31) and have negligible Stokes $\mathcal{V}$ as measured by interferometric-ALMA (Goddi et al in prep 2020), making them ideal for D-term calibration and polarimetric imaging. Their total-intensity structure is, however, more uncertain and more complex than M87. Both sources are blazar sources with bright extended jets (e.g., Wills & Wills 1981; Preston et al. 1989; Shen et al. 1997; Bower & Backer 1998; Healey et al. 2008), and their current EHT coverage may not capture the complexity of these jet sources. Nevertheless, their weak polarization allows for better D-term estimates despite uncertainty in the modeling of the structure.

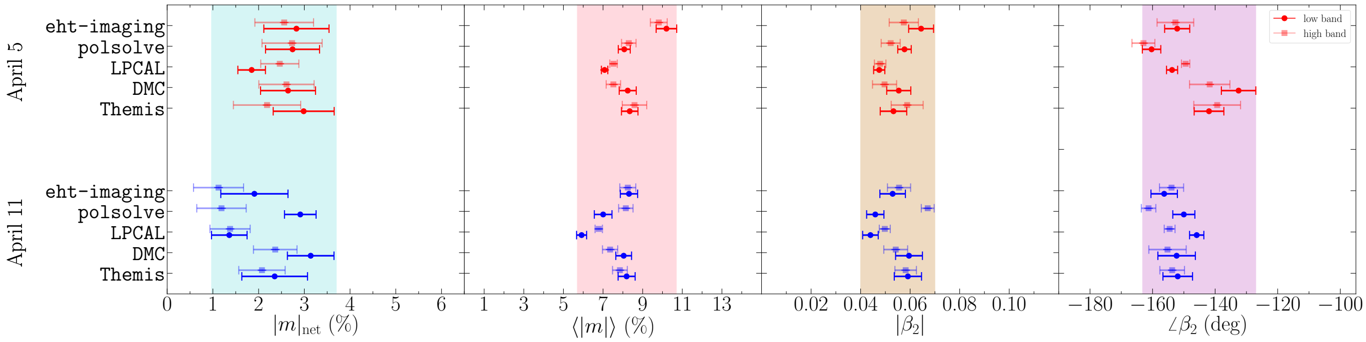

Figure 29. Comparison of high and low band results for the key quantities used in Paper VIII on April 5 and 11. The low band results are indicated by circular markers and the high band results are plotted in a lighter color with square markers. The low band results were presented in the main text in Figure 9. The vertical bands indicate the derived parameter ranges from the low-band results presented in Table 2.

Following the same methodology as M87, we generate synthetic data to optimize imaging and calibration parameters for all methods based on J1924–2914 and NRAO 530 low-band coverage. We use the same six ring-like synthetic models as M87 (see Section 4.3) and add a seventh model constructed with ten Gaussian sources of varying total and polarization intensity, with some polarization structure offset from Stokes $\mathcal{I}$. This seventh data set is designed to mimic basic structure seen in the polarimetric images of the two calibrators, for which the final images will be presented in future individual publications (S. Issaoun et al. *in prep.*, S. Jorstad et al. *in prep.*). We generate seven synthetic EHT observations for each source using their best EHT (u, v) coverage, April 11 and April 7 for J1924–2914 and NRAO 530 respectively. Parameter surveys are carried out for each method probing the same parameter space as for M87, and fiducial sets were selected with the same selection metrics, see Appendix G.

In Figures 32 and 33 we present the set of fiducial images from synthetic reconstructions using J1924–2914 and NRAO 530 best-day low-band coverage respectively. In each panel, the correlations between the ground truth and reconstructed Stokes $\mathcal{I}$ and linear polarization P images are provided. Consistent with the results with M87 coverage, the Stokes $\mathcal{I}$ correlations are high for all models regardless of method and coverage, and P correlations seem to worsen for models with complex polarization structure or high polarization.

In Figure 34 we compare the recovered leakage D-terms to the ground truth D-terms for the synthetic data sets with coverage from J1914–2914 (top) and NRAO 530 (bottom) and each method. Similarly to the M87 results, PV and SPT have the largest standard deviations for all methods. Their large deviations stem from all methods having difficulty recovering D-terms for models with no strong polarization substructure due to them being isolated stations with only long baselines. Overall, deviations of the D-terms measured via the L_1 norm (and its standard deviation) for the calibrators are comparable to those for M87 for all methods, but the standard deviation on each D-term estimate is noticeably wider for all stations, indicating that while overall image recovery is similar, the coverage differences between the M87 and the calibrator synthetic data do add uncertainty in the D-term recovery.

Finally, we estimate LMT, SMT and PV D-terms via polarimetric imaging of J1924–2914 and NRAO 530. The polarimetric images of these two calibrators will be presented in forthcoming publications (S. Issaoun et al. *in prep.*, S. Jorstad et al. *in prep.*). Here, in Figure 35, we show that D-terms of LMT, SMT, and PV estimated by imaging the calibrators roughly agree with those of M87. We note that a better agreement is obtained between M87 and J1924–2914 D-terms compared to M87 and NRAO 530. The calibrators have sparser (u, v) coverage (fewer scans), narrower field rotation range, and more complex Stokes $\mathcal{I}$ (extended structure and higher noise level) and polarimetric images compared to M87, which impact D-term estimation quality. Given these additional complexities, we argue that the calibrator D-terms are consistent with those of M87 (the D-term consistency within 2–3% is expected for the calibrators, see also Appendix K) and that M87 itself is the best source for polarimetric leakage calibration.

Furthermore, while imaging calibrators we found that the quality of Stokes $\mathcal{I}$ is critical for calibration. Both NRAO 530 and J1924–2914, being blazar sources, have complex jet structure that is not fully recovered with the current EHT coverage, and thus our Stokes $\mathcal{I}$ reconstructions have larger uncertainties and noise levels that those of M87, due to unconstrained flux density on large scales not sampled by our array configuration. Stokes $\mathcal{I}$ image assumptions in the polarimetric imaging and calibration methods can affect the results, for example the self-similarity assumption employed for CLEAN reconstructions in our sub-component methods, see Appendix K.

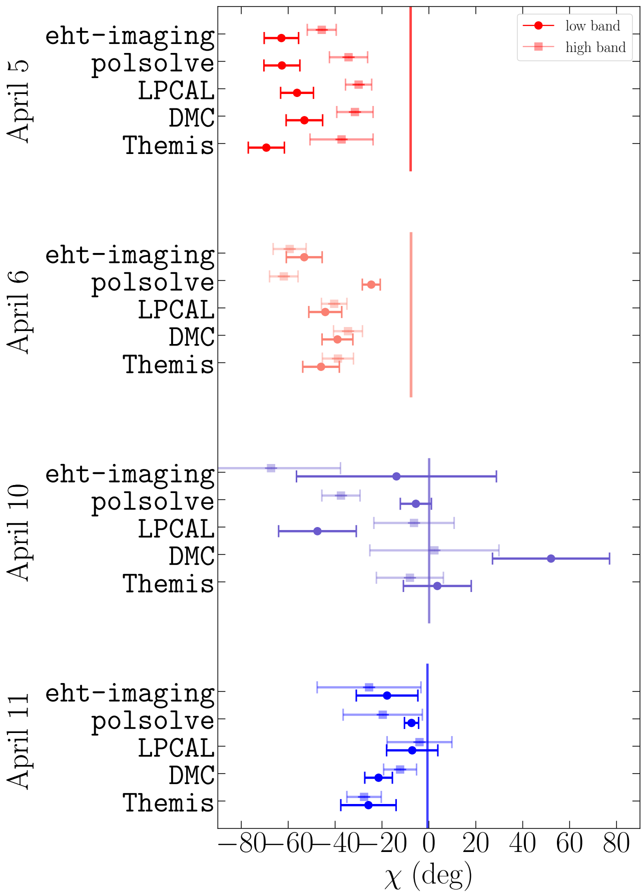

Figure 30. EHT net EVPA integrated within 120 μas on all four days for both high and low bands. The low band results are indicated by circular markers and the high band results are plotted in a lighter color with square markers. Vertical lines mark the ALMA-only EVPA measurements from Goddi et al. (2019), measured on arcsecond scales. The low band results for April 5 and 11 were already shown in Figure 10. The image-integrated EVPAs are consistent between high and low band on all days except April 5, where all methods show an offset of $\sim 20^{\circ}$. Note that the image reconstructions and EVPA measurements on April 10 are more uncertain due to the poor (u, v) coverage.

K. VALIDATION OF THE SIMILARITY APPROXIMATION IN CLEAN ALGORITHMS

The D-term estimation using the M87 data with `polsolve` and `LPCAL` reported in Section 4.2 are based on the similarity approximation. The Stokes $\mathcal{I}$ CLEAN models are divided into many sub-models to give a high degree of freedom for modelling the source's linear polarization structures for both software packages. Nevertheless, the complex linear polarization structures of M87 (Figure 6) may not be perfectly modelled with this approximation and this could be a source of uncertainties in the D-term estimation.

We investigate the effect of the similarity approximation by using the instrumental polarization self-calibration mode in `GPCAL` (see Section C.1). We ran `GPCAL` on the M87 data on April 11. The Stokes $\mathcal{I}$ CLEAN components are divided into 15 sub-models for initial D-term estimation using the similarity approximation. The D-terms of ALMA, APEX, and SMA are fixed to be zero for fitting as they were already calibrated using the intra-site baselines (Section 4.2). The intra-site baselines are flagged since the limited field-of-view of the EHT does not allow us to properly model the source structures observed on the short baselines. Instrumental polarization self-calibration was performed with ten iterations by employing `Difmap` for producing the Stokes Q and U images with CLEAN.

The left panel of Figure 36 compares the D-terms obtained by `GPCAL` with the fiducial D-terms on the same day, i.e., the average of `eht-imaging`, `polsolve`, and `LPCAL` results (Figure 2). Both the results using (i) the similarity approximation only (open squares) and (ii) the similarity approximation followed by ten iterations of instrumental polarization self-calibration (filled circles) are shown. Both of them show a good consistency with the fiducial D-terms with the L_1 norms of $\approx 1.0 - 1.2\%$, which is similar to the deviations between values of different pipelines seen in Figure 2. This result indicates that the D-terms obtained by `polsolve` and `LPCAL` using the M87 data are robust against the similarity approximation.

However, this may not be the case for the calibrators. We run `GPCAL` on the J1924–2914 data on April 11 and NRAO 530 data on April 7. Similar parameters to the M87 data analysis are used. The results are shown in the middle and right panels of Figure 36. The D-terms obtained with instrumental polarization self-calibration are more consistent with the fiducial D-terms than those obtained with the similarity approximation only. The L_1 norms improve from 9.2% to 1.4% and 2.9% to 2.0% for J1924–2914 and NRAO 530, respectively, with the instrumental polarization self-calibration. This result indicates that the linearly polarized structures of the calibrators are complex and cannot be easily modeled with the similarity approximation.

REFERENCES

Akiyama, K., Kuramochi, K., Ikeda, S., et al. 2017a, ApJ, 838, 1

Akiyama, K., Ikeda, S., Pleau, M., et al. 2017b, AJ, 153, 159

Baade, W. 1956, The Astrophysical Journal, 123, 550

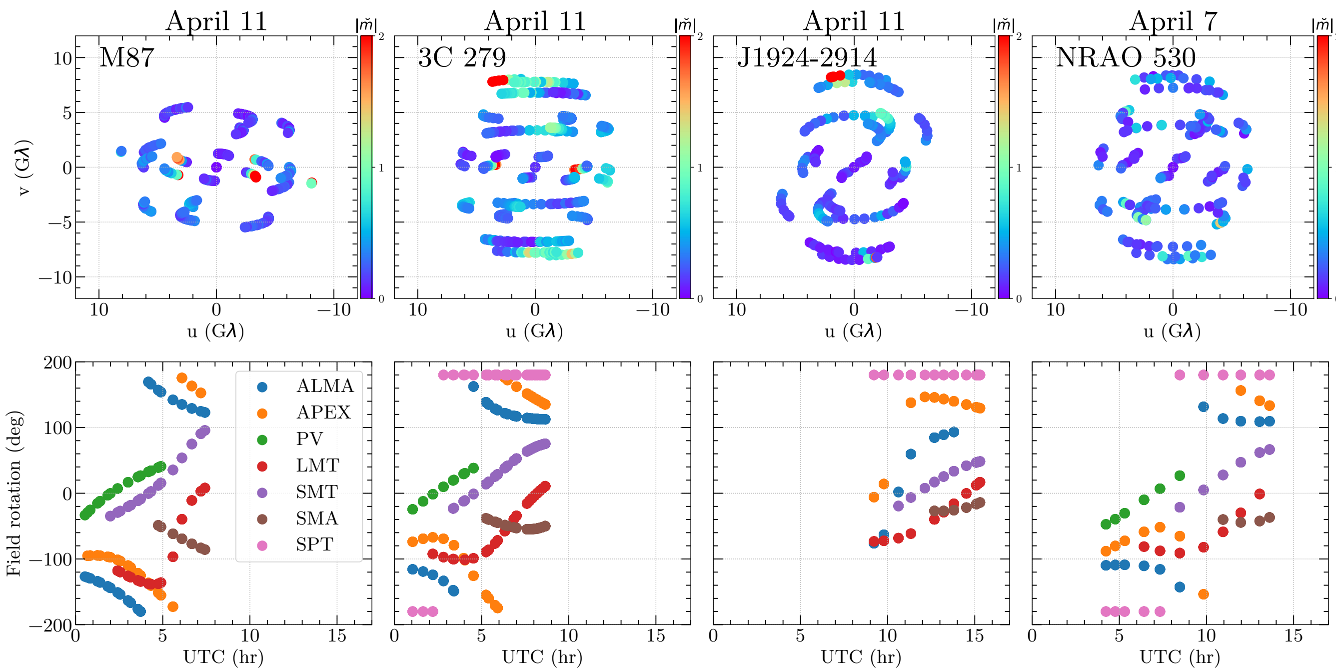

Figure 31. Top panels: Comparison of the polarization, (u, v) coverage and field rotation angle coverage of the main target and the calibrators. April 11 is shown for M87, 3C 279, J1924–2914 while April 7 is shown for NRAO 530. Color scales indicates fractional polarization amplitude $|\breve{m}|$ in the range from 0 to 2. Bottom panels: sources field rotation angle ϕ for each station as a function of time. The figure is analogous to Figure 1 for M87 on April 5, 6, 10 and 11.

Bird, S., Harris, W. E., Blakeslee, J. P., & Flynn, C. 2010, A&A, 524, A71

Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, ApJ, 894, 31

Blackburn, L., Chan, C.-k., Crew, G. B., et al. 2019, ApJ, 882, 23

Blakeslee, J. P., Jordán, A., Mei, S., et al. 2009, ApJ, 694, 556

Boccardi, B., Krichbaum, T. P., Ros, E., & Zensus, J. A. 2017, A&A Rv, 25, 4

Bower, G. C., & Backer, D. C. 1998, ApJL, 507, L117

Briggs, D. S. 1995, in Bulletin of the American Astronomical Society, Vol. 27, American Astronomical Society Meeting Abstracts, 1444

Broderick, A. E., & Pesce, D. W. 2020, ApJ, 904, 126

Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020a, ApJ, 898, 9

Broderick, A. E., Gold, R., Karami, M., et al. 2020b, ApJ, 897, 139

Burn, B. 1966, Monthly Notices of the Royal Astronomical Society, 133, 67

Cabral, B., & Leedom, L. C. 1993, in Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '93 (New York, NY, USA: Association for Computing Machinery), 263–270. https://doi.org/10.1145/166117.166151

Cantiello, M., Blakeslee, J. P., Ferrarese, L., et al. 2018, The Astrophysical Journal, 856, 126

Capetti, A., Macchetto, F., Sparks, W., & Biretta, J. 1997, Astronomy and Astrophysics, 317, 637

Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, Journal of Statistical Software, 76, 1

Chael, A., Narayan, R., & Johnson, M. D. 2019, MNRAS, 486, 2873

Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, ApJ, 857, 23

Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, ApJ, 829, 11

Cotton, W. D. 1993, AJ, 106, 1241

Cotton, W. D. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 82, Very Long Baseline Interferometry and the VLBA, ed. J. A. Zensus, P. J. Diamond, & P. J. Napier, 289

Curtis, H. D. 1918, Publications of Lick Observatory, 13, 9

Deller, A. T., Brisken, W. F., Phillips, C. J., et al. 2011, PASP, 123, 275

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJL, 875, L1, (Paper I)

—. 2019b, ApJL, 875, L2, (Paper II)

—. 2019c, ApJL, 875, L3, (Paper III)

—. 2019d, ApJL, 875, L4, (Paper IV)

—. 2019e, ApJL, 875, L5, (Paper V)

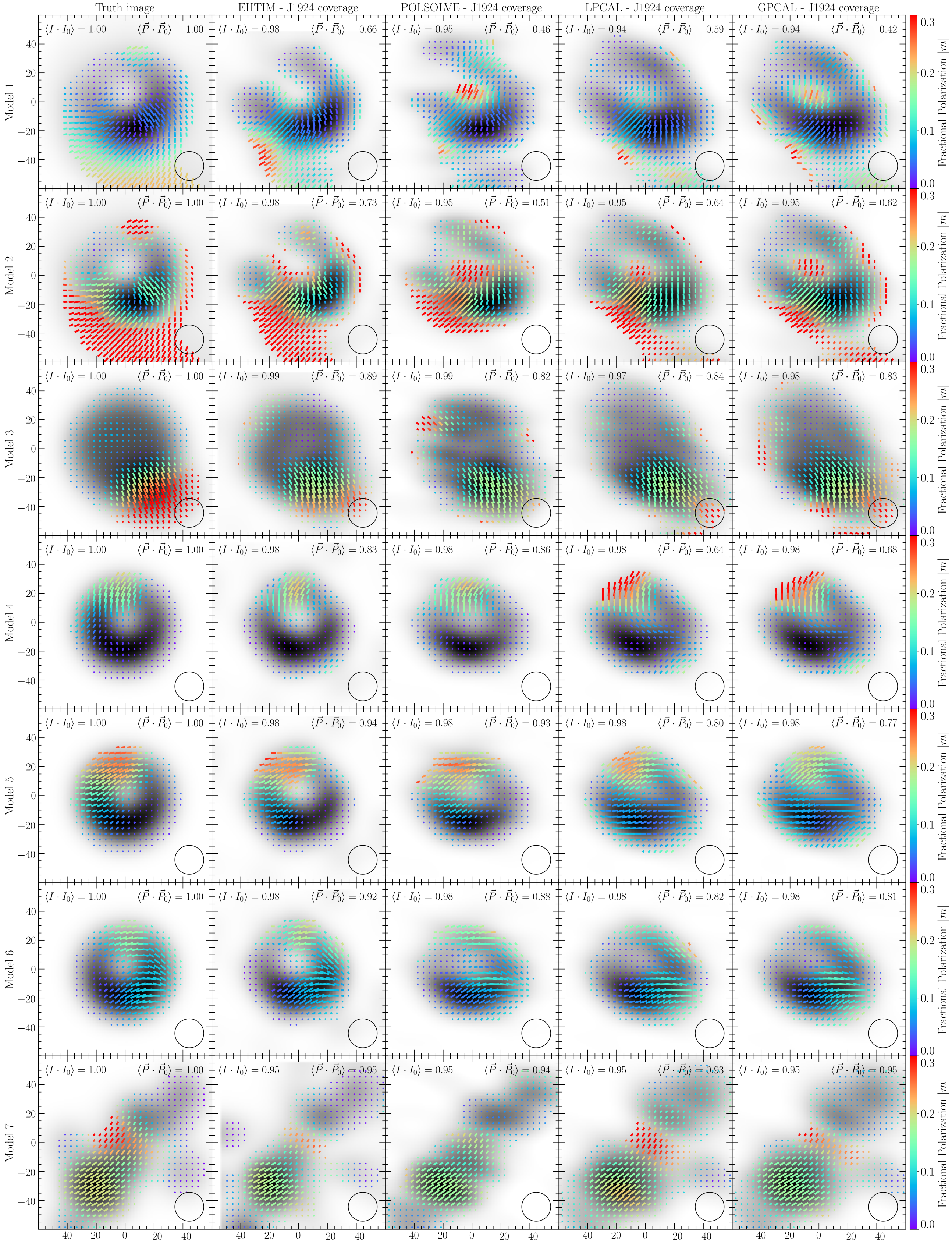

Figure 32. Fiducial images from synthetic data model reconstructions using J1924–2914 low band (u, v) coverage on April 11. Polarization tick length reflects total polarization, while color reflects fractional polarization from 0 to 0.3. Normalized overlap is calculated against the respective ground truth image, and for the case of total intensity is mean-subtracted.

—. 2019f, ApJL, 875, L6, (Paper VI)

—. 2020, in prep., (Paper VIII)

Gebhardt, K., Adams, J., Richstone, D., et al. 2011, The Astrophysical Journal, 729, 119

Goddi, C., Martí-Vidal, I., Messias, H., et al. 2019, PASP, 131, 075003

Goddi et al in prep, C. e. a. 2020, X

Greisen , E. W. 2003, in Astrophysics and Space Science Library, Vol. 285, Information Handling in Astronomy - Historical Vistas, ed. A. Heck, 109

Hada, K., Kino, M., Doi, A., et al. 2016, The Astrophysical Journal, 817, 131

Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137

Healey, S. E., Romani, R. W., Cotter, G., et al. 2008, ApJS, 175, 97

Hiltner, W. 1959, The Astrophysical Journal, 130, 340

Högbom, J. A. 1974, A&AS, 15, 417

Holdaway, M. A., & Wardle, J. F. C. 1990, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 1351, Digital Image Synthesis and Inverse Optics, ed. A. F. Gmitro, P. S. Idell, & I. J. Lahaie, 714–724

Homan, D. C., & Lister, M. 2006, The Astronomical Journal, 131, 1262

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90

Issaoun, S., Folkers, T. W., Blackburn, L., et al. 2017, EHT Memo Series, 2017-CE-02

Janssen, M., Issaoun, S., Blackburn, L., et al. 2019a, EHT Memo Series, 2019-CE-01

Janssen, M., Goddi, C., van Bemmel, I. M., et al. 2019b, A&A, 626, A75

Johnson, M. D., Fish, V. L., Doeleman, S. S., et al. 2015, Science, 350, 1242

Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, , . http://www.scipy.org/

Jones, R. C. 1941, Journal of the Optical Society of America (1917-1983), 31, 488

Kettenis, M., van Langevelde, H. J., Reynolds, C., & Cotton, B. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 497

Kim, J.-Y., Krichbaum, T. P., Broderick, A. E., et al. 2020, A&A, 640, A69

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt (IOS Press), 87 – 90

Kravchenko, E., Giroletti, M., Hada, K., et al. 2020, Astronomy & Astrophysics Letters, 637, L6

Kuo, C. Y., Asada, K., Rao, R., et al. 2014, ApJL, 783, L33

Leppänen, K. J., Zensus, J. A., & Diamond, P. J. 1995, AJ, 110, 2479

Li, Y.-P., Yuan, F., & Xie, F.-G. 2016, The Astrophysical Journal, 830, 78

Marrone, D. P. 2006, PhD thesis, Harvard University

Marrone, D. P., & Rao, R. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7020, Millimeter and Submillimeter Detectors and Instrumentation for Astronomy IV, ed. W. D. Duncan, W. S. Holland, S. Withington, & J. Zmuidzinas, 70202B

Martí-Vidal, I., Roy, A., Conway, J., & Zensus, A. J. 2016, A&A, 587, A143

Martí-Vidal, I., Krichbaum, T. P., Marscher, A., et al. 2012, A&A, 542, A107

Matthews, L. D., Crew, G. B., Doeleman, S. S., et al. 2018, PASP, 130, 015002

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51 – 56

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127

Owen, F. N., Eilek, J. A., & Keel, W. C. 1990, ApJ, 362, 449

Owen, F. N., Hardee, P. E., & Cornwell, T. 1989, The Astrophysical Journal, 340, 698

Palumbo, D. C. M., Wong, G. N., & Prather, B. S. 2020, ApJ, 894, 156

Park, J., Byun, D.-Y., Asada, K., & Yun, Y. 2021, ApJ, 906, 85

Park, J., Hada, K., Kino, M., et al. 2019, ApJ, 871, 257

Preston, R. A., Jauncey, D. L., Meier, D. L., et al. 1989, AJ, 98, 1

Rudin, L. I., Osher, S., & Fatemi, E. 1992, Physica D Nonlinear Phenomena, 60, 259

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55. https://doi.org/10.7717/peerj-cs.55

Shen, Z. Q., Wan, T. S., Moran, J. M., et al. 1997, AJ, 114, 1999

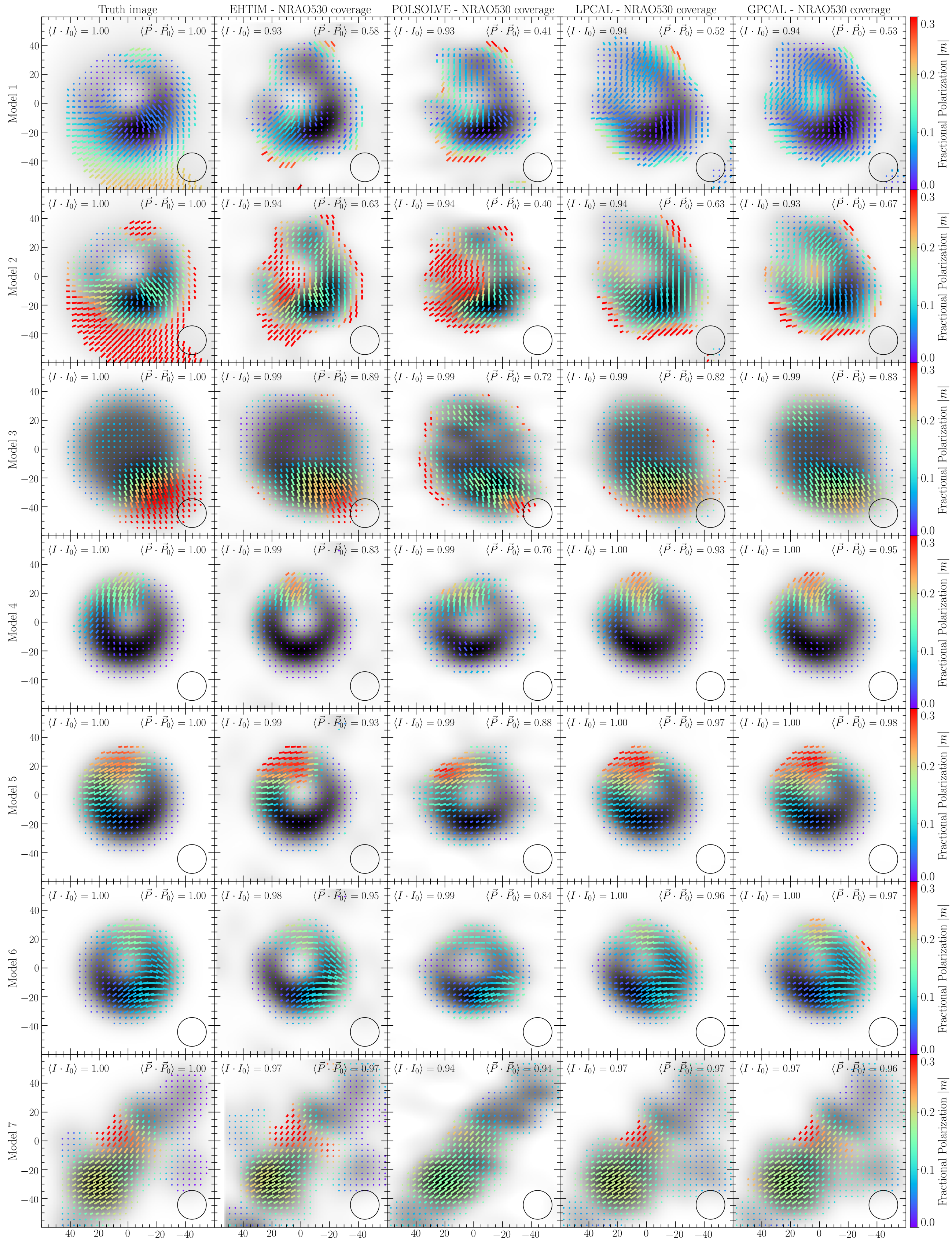

Figure 33. Fiducial images from synthetic data model reconstructions using NRAO 530 low band (u, v) coverage on April 7. Polarization tick length reflects total polarization, while color reflects fractional polarization from 0 to 0.3. Normalized overlap is calculated against the respective ground truth image, and for the case of total intensity is mean-subtracted.

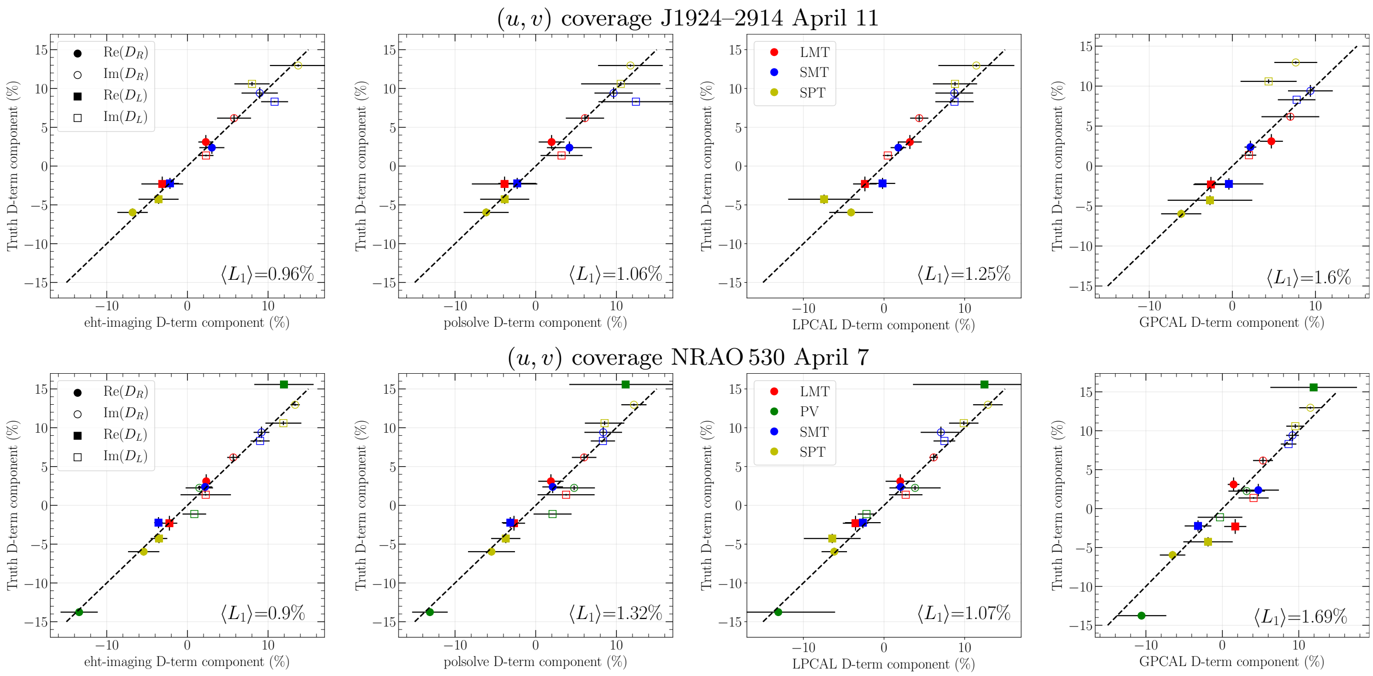

Figure 34. D-terms for LMT, SMT, PV and SPT derived from synthetic datasets. A comparison of estimates to ground truth values is shown per software (`eht-imaging`, `polsolve`, `LPCAL` and `GPCAL` results are shown in first through fourth columns, respectively) and per (u, v) coverage of the real observations (results for (u, v) coverage of J1924–2914 on April 11 and the (u, v) coverage of NRAO 530 on April 7 are shown in the top and bottom panels, respectively). Each data point corresponds to an average and standard deviation for each D-term estimate derived from synthetic data sets 1-7. The norm $L_1 \equiv |D - D_{\rm Truth}|$ is averaged over left, right, real, and imaginary components of the D-terms and over all shown EHT stations.

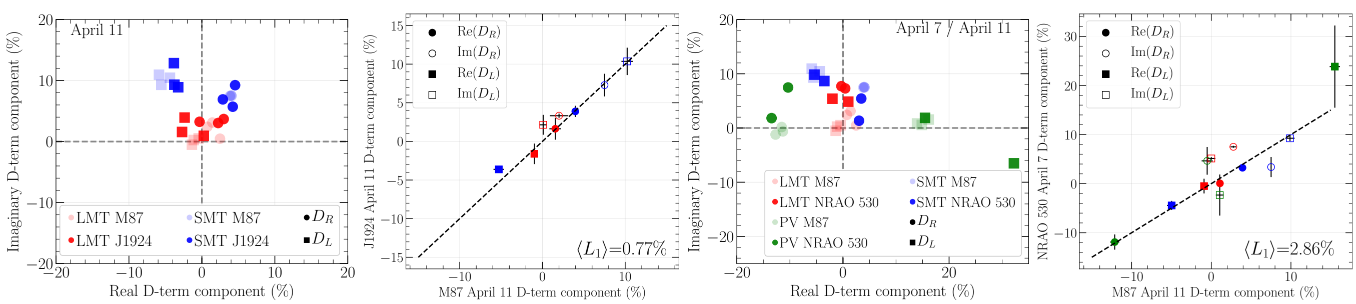

Figure 35. Comparison of fiducial D-terms for the telescopes LMT, SMT, and PV estimated from M87 (April 11), J1924–2914 (April 11) and NRAO 530 (April 7) low band data sets using `eht-imaging`, `polsolve`, and `GPCAL` tools. In the first and third panel the M87 D-terms are depicted with lighter symbols while heavier symbols mark the calibrator D-terms. In the correlation plots shown in the second and fourth panels, the D-terms for M87 and J1924–2914/NRAO 530 are averaged over different tools. LMT and SMT D-terms derived from J1924–2914 are found to be highly consistent with those from M87. The D-terms derived from NRAO 530 imaging on average show larger deviation from M87 D-terms; in particular, **the** PV D-terms estimated by `eht-imaging` show the largest deviation from all other estimates.

Shepherd, M. 2011, Difmap: Synthesis Imaging of Visibility Data, Astrophysics Source Code Library, , , ascl:1103.001

Smirnov, O. M. 2011, A&A, 527, A106

Steel, S., Wielgus, M., Blackburn, L., Issaoun, S., & Johnson, M. 2019, EHT Memo Series, 2019-CE-03

Syed, S., Bouchard-Côté, A., Deligiannidis, G., & Doucet, A. 2019, arXiv e-prints, arXiv:1905.02939

Tange, O. 2011, ;login: The USENIX Magazine, 36, 42. http://www.gnu.org/s/parallel

The Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123

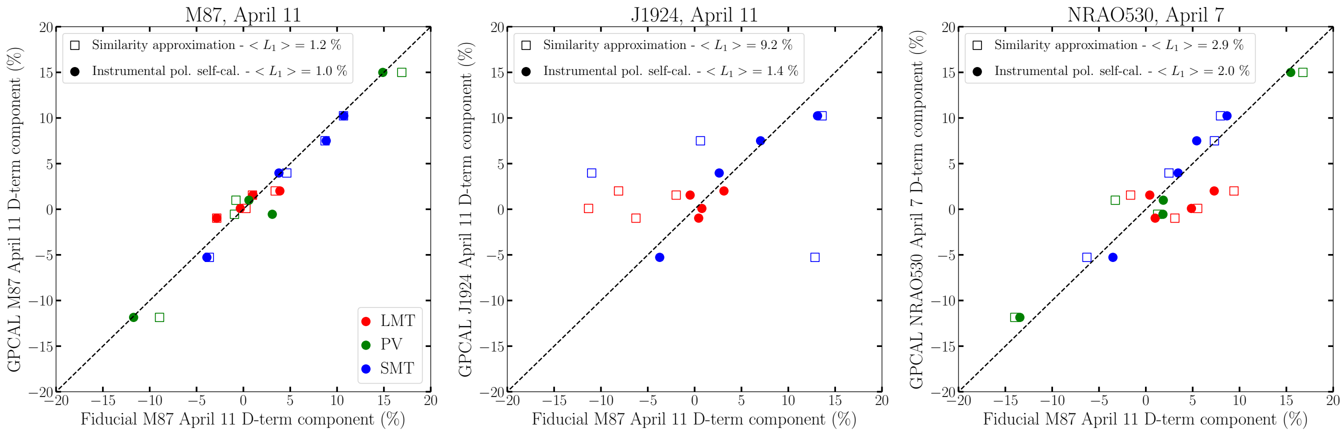

Figure 36. Comparison of D-terms estimated with `GPCAL` with and without using instrumental polarization self-calibration. The average of the D-terms estimated with `eht-imaging`, `polsolve`, and `LPCAL` on the M87 data on April 11 are shown on the $x-$axis. The `GPCAL` results with (filled circles) and without (open squares) using instrumental polarization self-calibration are shown on the $y-$axis (see text for more details). The left, middle, and right panels show the results of M87 on April 11, J1924–2914 on April 11, and NRAO 530 on April 7, respectively. The L_1 norms do not change much with instrumental polarization self-calibration for M87, while they are significantly improved for the calibrators, especially for J1924-2914. This indicates that the similarity approximation employed by `polsolve` and `LPCAL` for the D-term estimation from M87 (Section 4.2) is reasonable. The calibrators may have complex linear polarization structures and D-term estimation from those sources can be improved with instrumental polarization self-calibration.

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition (Springer International Publishing), doi:10.1007/978-3-319-44431-4

Thomson, R., Robinson, D., Tanvir, N., Mackay, C., & Boksenberg, A. 1995, Monthly Notices of the Royal Astronomical Society, 275, 921

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22

Walker, R. C., Hardee, P. E., Davies, F. B., Ly, C., & Junor, W. 2018, ApJ, 855, 128

Whitney, A. R., Cappallo, R., Aldrich, W., et al. 2004, Radio Science, 39, doi:10.1029/2002RS002820

Wielgus, M., Blackburn, L., Issaoun, S., Janssen, M., & Koay, J. 2019, EHT Memo Series, 2019-CE-02

Wills, D., & Wills, B. J. 1981, Nature, 289, 384

Zavala, R., & Taylor, G. 2002, The Astrophysical Journal Letters, 566, L9