LIFTING THE VEIL ON BLACK HOLES
APPROACHING THE EVENT HORIZON WITH HIGH-RESOLUTION IMAGING

SARA ISSAOUN
Lifting the Veil on Black Holes: 
Approaching the Event Horizon with High-resolution Imaging

Proefschrift

ter verkrijging van de graad van doctor 
aan de Radboud Universiteit Nijmegen 
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken, 
volgens besluit van het college van decanen 
in het openbaar te verdedigen op 
vrijdag 3 september 2021 
on 16:30 uur precies

doors

Sara Issaoun

geboren op 27 mei 1994 
te Boghni, Algerije
Promotor: Prof. dr. H.D.E. Falcke

Manuscriptcommissie: Prof. dr. B. Krishnan
Prof. dr. J.R. Hörandel
Prof. dr. G.A. Nelemans
Prof. dr. S.B. Markoff Universiteit van Amsterdam
Prof. dr. D.P. Marrone University of Arizona, Verenigde Staten
## Contents

1 Introduction ............................................. 1
   1.1 A Brief History of Black Holes ......................... 1
   1.2 Black Hole Accretion and Outflow ...................... 3
   1.3 Very Long Baseline Interferometry ..................... 7
   1.4 Interferometric Imaging .............................. 9
   1.5 A Tale Of Two Black Holes ......................... 10
       1.5.1 Sagittarius A* .................................. 11
       1.5.2 M87* ........................................... 13
   1.6 Interstellar Scattering .............................. 16
   1.7 This Thesis ......................................... 21

2 The Size, Shape, and Scattering of Sagittarius A* .......... 23
   2.1 Introduction ......................................... 24
   2.2 Background ......................................... 26
       2.2.1 Theoretical models for Sgr A* emission .......... 26
       2.2.2 Interstellar Scattering of Sgr A* ............... 26
   2.3 Observations and data reduction ...................... 28
   2.4 Imaging ............................................. 31
       2.4.1 Calibrators NRAO 530 and J1924–2914 ............ 31
       2.4.2 Sagittarius A* .................................. 32
   2.5 Results ............................................. 41
       2.5.1 Intrinsic source constraints from imaging ....... 41
       2.5.2 Intrinsic source constraints from closure phases . 42
   2.6 Discussion ......................................... 44
       2.6.1 Constraints on the Refractive Scattering of Sgr A* . 44
       2.6.2 Constraints on accretion flow and jet models .... 46
   2.7 Summary ............................................. 52

3 VLBI imaging of black holes via second moment regularization .... 55
   3.1 Introduction ......................................... 55
   3.2 Background ......................................... 57
3.3 Method .................................................. 61
  3.3.1 Second moment regularization .......................... 62
  3.3.2 Assumptions ........................................... 62
3.4 Demonstration ............................................. 65
  3.4.1 Imaging with complementary size constraints ....... 67
  3.4.2 Dependence of reconstructed images on assumed size . 69
  3.4.3 Imaging without complementary size constraints .... 72
3.5 Applications ............................................. 72
  3.5.1 Scattering mitigation .................................. 73
  3.5.2 Dynamical imaging ..................................... 76
3.6 Summary .................................................. 76
3.A Properties of the visibility function ....................... 77
  3.A.1 Visibility derivatives and image moments .......... 77
  3.A.2 Image principal axes and visibility curvature ....... 78
3.B Implementation via gradient descent ....................... 79
  3.B.1 Pixel-based derivation of principal axes .......... 79
  3.B.2 Gradient Descent Implementation ..................... 80
4 Persistent Non-Gaussian Structure in the Image of Sagittarius A* 83
  4.1 Introduction ............................................. 84
  4.2 Observations and data reduction ......................... 87
  4.3 Results ................................................ 89
    4.3.1 Imaging the Calibrator J1924–2914 .................. 90
    4.3.2 Calibrating Sagittarius A* Visibility Amplitudes 92
    4.3.3 Final Sgr A* Visibility Amplitudes ................ 96
  4.4 Discussion ............................................. 97
    4.4.1 Long-Wavelength Constraints on Inner Scale and Intrinsic Size 99
    4.4.2 Joint Modeling of Scattering and Source Parameters 99
    4.4.3 Constraints on the Power Spectrum .................. 102
  4.5 Summary ................................................ 104
5 EHT Data Processing and Calibration .......................... 107
  5.1 Introduction ............................................. 108
  5.2 Observations ............................................ 110
  5.3 Data Flow .............................................. 112
  5.4 Correlation .............................................. 114
  5.5 Fringe Detection ......................................... 116
    5.5.1 HOPS Pipeline ...................................... 116
    5.5.2 CASA Pipeline ..................................... 119
    5.5.3 AIPS Pipeline ..................................... 121
  5.6 Flux Density Calibration .................................. 123
    5.6.1 A Priori Amplitude Calibration ....................... 123
    5.6.2 Network Calibration .................................. 133
## Contents

- **5.7 Final Data Products** ........................................... 135
  - 5.7.1 Data Release Specification ............................. 135
  - 5.7.2 Closure Quantities ..................................... 137
  - 5.7.3 Data Features ........................................... 138
- **5.8 Data Validation and Systematics** .......................... 142
  - 5.8.1 Fringe Validation ....................................... 143
  - 5.8.2 Thermal Error Consistency ............................ 143
  - 5.8.3 Temporal Coherence After Calibration ................. 146
  - 5.8.4 Intra-Pipeline Validation ............................. 149
  - 5.8.5 Inter-Pipeline Consistency ............................ 154
- **5.9 Conclusions** ................................................ 157
- **5.A Site and Data Issues** ...................................... 159
  - 5.A.1 Issues Requiring Mitigation ............................ 159
  - 5.A.2 Issues Not Addressed During Processing ............... 161
  - 5.A.3 Issues at Correlation .................................... 162
- **6 Imaging a Black Hole with the EHT** ........................ 163
  - **6.1 Total Intensity Imaging of M87** ....................... 166
    - 6.1.1 Observations and Data Processing ..................... 166
    - 6.1.2 First M87 Images from Blind Imaging ............... 168
    - 6.1.3 Parameter Surveys and Final Images .................. 176
    - 6.1.4 Image Validation with 3C279 .......................... 181
  - **6.2 Polarimetric Imaging of M87** ........................... 185
    - 6.2.1 Basic definitions .................................... 185
    - 6.2.2 EHT 2017 Polarimetric Data ......................... 186
    - 6.2.3 Methods for Polarimetric Imaging and Leakage Calibration ................. 189
    - 6.2.4 Software Conventions and Organization ............. 190
    - 6.2.5 Leakage and Gain Calibration Strategy .............. 192
    - 6.2.6 Preliminary Polarimetric Results .................... 195
    - 6.2.7 Parameter Surveys and Validation on Synthetic Data ...... 196
    - 6.2.8 Fiducial Polarimetric Images of M87 ................. 200
  - **6.A Multi-Day Gain Comparisons with 3C279** ............ 203
  - **6.B Polarimetric Data Issues** ............................... 203
    - 6.B.1 Instrumental polarization of ALMA in VLBI mode . 203
    - 6.B.2 Instrumental polarization of the LMT ............... 206
    - 6.B.3 Instrumental polarization of the SMA ............... 206
    - 6.B.4 Instrumental polarization of the JCMT .............. 207
  - **6.C LMT, SMT and PV D-terms using calibrator data: synthetic data tests, expected uncertainties and convergence with M87 results** ................. 207
- **A A conceptual overview of single-dish absolute amplitude calibration** 213
  - **A.1 Introduction to standard single-dish $T_{sys}$ calibration techniques** ................. 215
  - **A.1.1 The antenna-based system-equivalent flux densities (SEFDs)** ................. 215
1.1 A Brief History of Black Holes

In 1784, Reverend John Michell pondered on the possibility of having a star so small that its escape velocity would be larger than the speed of light. In Newtonian physics, the escape velocity of a star is:

$$v_{\text{esc}} = \sqrt{\frac{GM_\star}{R_\star}},$$

(1.1)

where $G$ is the Newtonian gravitational constant, and $M_\star$ and $R_\star$ are the mass and radius of the star, respectively. For a star of a given mass, reducing the size of the star (its radius) increases the escape velocity. This thought experiment showed that once the escape velocity exceeds the speed of light, photons are trapped on the surface of the star, and the star will appear dark to a distant observer. He coined the term ‘dark star’ to name this theoretical object [Michell, 1784].

Nearly two centuries later, in 1915, Albert Einstein publishes his general theory of relativity (GR), where gravity is not a force but a curvature of the fabric of spacetime. His theory predicts that light is affected by gravity, and massive objects are able to bend the path of light [Einstein, 1915]. This prediction was later measured with the solar eclipse experiment carried out by Arthur Eddington and his team in 1919. In 1916, in the trenches of the First World War, Karl Schwarzschild discovers the first non-trivial analytical vacuum solution to Einstein’s equations of GR, resulting in a static spherically symmetric spacetime, with a singularity at its center. The singularity is enclosed in a causally disconnected region of spacetime circumscribed by a boundary surface that depends on a particular radius, the ’Schwarzschild’ radius [Schwarzschild, 1916]:

$$R_{\text{Sch}} = \frac{2GM}{c^2}.$$  

(1.2)

This mathematical boundary is what we call the ‘event horizon’. Anything crossing the event horizon would require an escape velocity greater than the speed of light to emerge, and so only
darkness remains [Finkelstein, 1958]: a black hole (although the term ‘black hole’ was not popularized until the 1960s). In the same year, David Hilbert publishes his lectures on ‘The Foundations of Physics’. He becomes the first to calculate the appearance of a Schwarzschild black hole as seen by a distant observer, a dark region with a diameter of $\sqrt{27}R_{\text{Sch}}$, surrounded by bending rays of light [see Sauer & Majer, 2009]. Max von Laue shortly followed in 1921 [von Laue, 1921]. At this point, black holes are a mathematical occurrence of Einstein’s theory of GR, but have yet to be linked to real astrophysical phenomena as it is unclear how such objects could form.

In 1931, Subrahmanyan Chandrasekhar discovers that the end points of solar-mass stars, white dwarfs, have a maximum mass [Chandrasekhar, 1931]. More massive stars must thus have a different end point. When a massive star exhausts its fuel at the end of its lifetime, the thermal pressure from nuclear fusion can no longer balance its gravity, its electron-degenerate core overcomes degeneracy pressure and its core collapses into a neutron core in a supernova explosion of the outer material. The core collapse leads to a neutron core being packed in such a small volume that it becomes a stellar mass black hole if its mass is beyond the Tolman-Oppenheimer-Volkov limit [Tolman, 1939; Oppenheimer & Volkoff, 1939]. This scenario is the first time that black holes are considered physical objects that could exist in the Universe. Such black holes are now commonly found in X-ray binary systems, where the accreted material from a companion star onto the black hole creates bright X-ray emission [e.g., Cygnus X-1; Bolton, 1972; Webster & Murdin, 1972]. More massive stellar-mass black holes ($\sim 10 - 100M_\odot$) have also recently been observed via the detection of gravitational waves by the LIGO/VIRGO collaboration [e.g., Abbott et al., 2016]. On a much larger scale, supermassive black holes ($\sim 10^5 - 10^{10}M_\odot$) reside at the center of most large galaxies [e.g., Salpeter, 1964; Lynden-Bell, 1969]. Cygnus A, first discovered in the radio sky by Reber [1944], was the first source to be proposed as a supermassive black hole candidate. Smith [1951] finds that the source is of extragalactic origin and extremely luminous. In the 1960s, multiple luminous radio sources, like the famous 3C273, are discovered that are all of extragalactic origin, and are dubbed ‘quasars’ [QSR for ‘quasi-stellar radio source’: e.g., Schmidt, 1963; Greenstein, 1963; Schmitt, 1968]. Models are proposed to explain the high luminosity of these objects, powered by accretion and toroidal magnetic fields around a supermassive central engine (see 1.2). These black holes are believed to grow from core collapse of local clusters, galaxy mergers, and accretion of surrounding material over millions to billions of years [e.g., Ferrarese & Merritt, 2000; Gebhardt et al., 2000].

In 1963, Roy Kerr discovers the solution to Einstein’s equations for a rotating black hole and provides an exact description of all astrophysical black holes: the ‘no hair’ theorem [Kerr, 1963]. The no-hair theorem stipulates that any black hole can be entirely described by three quantities: its mass, its spin, and its charge [for which the solution was found by Ezra Newman; Newman et al., 1965]. In 1973, James Bardeen shows that the shape of a black hole as seen by a distant observer changes with its spin: a spinning black hole’s appearance will deviate slightly from circularity, unlike a non-rotating Schwarzschild black hole [see Figure 1.1; Bardeen, 1973]. Observing the shape deviation from circularity allows us to test fundamental properties of spacetime: while GR predicts deviations from circularity of only up to 4% with a maximally spinning black hole [Falcke et al., 2000; Takahashi, 2004; Johannsen & Psaltis, 2010], other theories of gravity predict larger deviations, and more exotic objects may also exhibit different signatures [e.g., Chirenti & Rezzolla, 2007; Bambi et al., 2009; Johannsen & Psaltis, 2010; Younsi et al.,
1.2 Black Hole Accretion and Outflow

The gravitational pull of a black hole causes infalling material to accrete onto it via various mechanisms. For idealized spherically symmetric accretion [Bondi-Hoyle accretion; Hoyle & Lyttleton, 1941; Bondi, 1952], the outward radiation force of the infalling gas must not exceed the gravi-
Figure 1.2: First computer generated simulated image of a lensed thin accretion disk around a spherically symmetric black hole from Luminet [1979].

Figure 1.3: Simulated visualizations of the inner accretion flow and shadow of Sagittarius A* with a black hole spin of 0.998 (top) and zero (bottom) from Falcke et al. [2000]. Left: models of the shadow at infinite resolution. Center and right: models of the shadow as observed with an idealized interferometric array at 0.6 mm and 1.3 mm respectively. Note these were made assuming a smaller black hole mass measurement for Sagittarius A* than the current agreed mass.
1.2 Black Hole Accretion and Outflow

tional pull of the accreting object. This limit where the two forces balance out is called the ‘Eddington luminosity’:

$$L_{\text{Edd}} = 1.3 \cdot 10^{38} \frac{M}{M_\odot} \text{ erg s}^{-1}. \quad (1.3)$$

The accretion rate of an object radiating at the Eddington luminosity is thus given by:

$$\dot{M}_{\text{Edd}} = \frac{\delta M}{\delta t} = \frac{L_{\text{Edd}}}{\eta c^2}, \quad (1.4)$$

where $\eta$ is the accretion efficiency of the process. However, spherically symmetric accretion is not very realistic, inflow of gas is not expected to be uniform but rather unsteady, and depends on movement of the surrounding gas from stellar winds or gas clouds or accreted stars. From the angular momentum conservation argument, infalling matter typically forms a disk around the accreting object instead [Weizsäcker, 1948; Lüst, 1952].

In 1973, Shakura and Sunyaev proposed a geometrically thin and optically thick accretion disk model called the $\alpha$-disk, with a viscosity prescription originating from gas turbulence [Shakura & Sunyaev, 1973; Novikov & Thorne, 1973]:

$$\nu_{\text{visc}} = \alpha c_s H, \quad (1.5)$$

where $\nu_{\text{visc}}$ is the kinematic viscosity, $\alpha$ is the tunable parameter, $c_s$ is the sound speed, and $H$ is the disk scale height. The viscosity in the disk leads to angular momentum transport and efficient accretion of the optically thick gas onto the central object. In this model, the temperature of the accreted gas is cold compared to the virial temperature of the flow, and the accretion energy is dissipated via blackbody radiation. The $\alpha$-disk model is thought to be a good representation of sources that accrete near Eddington luminosity [Frank et al., 2002].

A second class of models came about in the 80s, where the accretion disk is geometrically thick but optically thin. In these disks, called advection dominated accretion flows [ADAFs; Narayan & Yi, 1994; Narayan et al., 1998], the gas temperature is of the order of the virial temperature, and accretion energy is dissipated not via radiation but via viscous dissipation and heating of the gas, leading to low-luminosity sources. Black holes with much lower accretion rates are ‘starved’ via a class of ADAF models: a radiatively inefficient accretion flow [RIAF; Ichimaru, 1977; Rees et al., 1982; Narayan & Yi, 1994; Abramowicz et al., 1995]. In RIAFs, the electron and ion temperatures in the flow differ, as electrons cool much faster via radiative cooling, while ions cool at a slower rate via inefficient Coulomb interactions in an optically thin disk with collisionless plasma. The primary emission mechanism is thus synchrotron, where electrons gyrate at relativistic speeds around magnetic field lines and emit radiation that results in a power-law emission spectrum typically in the radio regime [see Figure 1.5; Rybicki & Lightman, 1979]. This model is believed to hold for low-luminosity active galactic nuclei (LLAGN), such as the supermassive black hole at the center of our Milky Way, Sagittarius A*, and that at the center of the galaxy M87.

For objects that accrete at or above Eddington rates, there is a variant of ADAF models where the disk is both geometrically and optically thick, called the slim disk [Begelman, 1979; Abramowicz et al., 1988]. For slim disks, radiation trapped inside the flow due to the optical thickness is advected toward the central object. Due to the high densities, increased collisions with the plasma create additional radiation pressure that puffs up the disk. This accretion state
is believed to hold for efficiently-accreting AGN and X-ray binaries [Frank et al., 2002]. In X-ray binaries, the disk thickness increases in the X-ray ‘hard state’ at the start of an X-ray outburst and becomes geometrically thin during the outburst, or ‘soft state’, back to its quiescent state [Fender et al., 2004]. In the case of very strong outbursts, the thickness of the disk can increase to the ‘slim disk’ ADAF regime at the start of an outburst.

In some systems, the inflow of material toward a central object and differential rotation between the magnetic fields and the disk material close to the black hole naturally lead to an arising outflow. This outflow can take the form of a disk wind or in the most powerful cases a relativistic jet. Relativistic jets are ubiquitous in the universe, piercing through entire galaxies and into the intergalactic medium, and are thought to have as a central engine the AGN at the center of the host galaxy. The first astrophysical jet was discovered in 1918 in the M87 galaxy by astronomer Heber Curtis [Curtis, 1918].

While the specifics of jet formation and propagation are yet to be fully understood, there are a number of proposed mechanisms to explain these processes. The two most commonly used in the community are the Blandford-Znajek and Blandford-Payne mechanisms [Blandford & Znajek, 1977; Blandford & Payne, 1982]. The Blandford-Znajek mechanism postulates that the rotational energy from a spinning black hole powers the jet. The accreting plasma causes a build up of magnetic fields near the black hole’s event horizon. Due to the strong pressure, the magnetic fields twist vertically along the black hole’s spin axis and form a helical structure that accelerates plasma near the black hole, emitting synchrotron radiation. In the Blandford-Payne mechanism, particles are accelerated in an outflow due to magnetic fields twisted by differential rotation within the accretion disk itself. As the jet propagates outward, observed knots and lobes in the jet structure are thought to be shock regions where the plasma from the jet collides with dense intergalactic medium, where particles can once again be accelerated and the jet can be re-collimated. In both mechanisms, magnetic fields play a central role in jet formation, and are also believed to maintain the jet in a thin stream as it propagates outward into the galactic

---

**Figure 1.4:** Image of the optical jet from M87 by the Hubble Space Telescope. This jet was the first astrophysical jet to be observed, by Curtis [1918] using the optical telescope of the Lick Observatory. Image credit: NASA and the Hubble Heritage Team (STScI/AURA).
Electrons moving at relativistic speeds emit synchrotron radiation as they gyrate around a magnetic field. The electric field of the emitted electromagnetic waves oscillates in the direction perpendicular to the direction of the local magnetic field line. Synchrotron radiation from relativistic electrons accelerated around magnetic field lines are commonly emitted at radio frequencies, thus radio observations of black hole systems are an ideal probe into accretion and jet launching mechanisms.

In addition, synchrotron radiation can be up to 75% polarized: the radio electromagnetic waves emitted by relativistic electrons have a specific direction of oscillation of their electric field perpendicular to the direction of travel [Rybicki & Lightman, 1979]. This direction is the normal of the direction of the local magnetic field line around which the electron is gyrating at the time of emission, as shown in the schematic in Figure 1.5. This direction of oscillation as measured by a distant observer is called the electric vector position angle (EVPA). Observing with instruments able to measure polarization intensity and direction further informs properties of the accretion flow and inference of magnetic field configuration, see Chapter 6.

1.3 Very Long Baseline Interferometry

The resulting interference pattern from the interaction of two waves informs the observer of the properties of the original waves: two waves in phase yield a constructive pattern while two waves out of phase yield a destructive pattern. With interferometers, we can measure the interference pattern of signals detected between two telescopes and infer information about spatially incoherent source structure on the sky.

Consider a point source in the far field of an interferometer: a source sufficiently far away that the incident electromagnetic waves arrive as plane waves on Earth. For an interferometer constructed of two antennas separated by a distance $D$ (or baseline, see Figure 1.6), there is a geometrical delay $\tau_g$ between the arrival times of the wavefront (arriving in the $\theta$ direction) at the two antennas:

$$\tau_g = \frac{D}{\sin \theta}.$$  

(1.6)

![Figure 1.5: Electrons moving at relativistic speeds emit synchrotron radiation as they gyrate around a magnetic field. The electric field of the emitted electromagnetic waves oscillates in the direction perpendicular to the direction of the local magnetic field line.](image-url)
In terms of frequency, the product of the incident waves at both antennas is proportional to what we call the ‘fringe function’

\[ F = 2 \sin(2\pi \nu t) \sin 2\pi \nu(t - \tau_g). \]  

(1.7)

Since the rate of variation of \( \theta \) depends on the Earth’s rotation, which is slow, we can neglect the \( \nu \tau_g \) term, and for an averaging period \( T >> 1/\nu \) we can further simplify the fringe function to

\[ F = \cos 2\pi \nu \tau_g = \cos \left( \frac{2\pi D x}{\lambda} \right), \]  

(1.8)

where \( \lambda \) is the observing wavelength and \( x = \sin \theta \). The widths of the fringes in the resulting fringe function pattern determine the resolution of the interferometer, and depend on \( \lambda \) and \( D \). The resolution \( \psi \) of a single telescope of diameter \( d \) is given by the Rayleigh criterion [Lord Rayleigh, 1879]:

\[ \psi = A \frac{\lambda}{d}, \]  

(1.9)

where \( A \) depends on the position of the first minimum in the diffraction pattern. For typical optical and radio telescopes, \( A \sim 1.22 \). For an interferometer, however, the resolution for an interferometer is not hindered by the diffraction limit and is given by:

\[ \psi = \frac{\lambda}{D}. \]  

(1.10)

An interferometer with a baseline \( D = d \) thus achieves a better resolution than a single telescope of diameter \( d \). As we increase the number of antennas in the interferometer, different resolutions and orientations are probed depending on the positioning of the antennas on the Earth in relation to each other.

For emission detected at a small angle \( \Delta \theta \) from the central source position, the fringe pattern is approximated to:

\[ F(l) \simeq \cos(2\pi ux), \]  

(1.11)

where \( x = \sin \Delta \theta \) and the spatial frequency \( u = D \cos \theta/\lambda \) is the projected baseline length between two antennas perpendicular to the source direction in wavelength units. For an extended
1.4 Interferometric Imaging

A (non-point-like) and spatially incoherent source with an intensity pattern $I$ on the sky, the interferometer response $r(u)$ is proportional to the convolution of the intensity function and the fringe pattern. In the Fourier domain, it is thus the product of the fringe pattern delta function and the visibility function $\mathcal{V}(u)$ (the Fourier transform of the intensity pattern):

$$r(u) = \frac{1}{2} [\delta(u + u_0) + \delta(u - u_0)] \mathcal{V}(u).$$  \hspace{1cm} (1.12)

The visibility function extended to two dimensions for a spatially incoherent source is given by the van Cittert-Zernike theorem [van Cittert, 1934; Zernike, 1938]:

$$\mathcal{V}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-2\pi i (ux + vy)} dxdy,$$  \hspace{1cm} (1.13)

where $u$ and $v$ are the projected baselines in the east-west and north-south directions respectively, and $x$ and $y$ are their respective cosine coordinates on the sky. Here we assume that $x$ and $y$ are small (the source is small on the sky) such that $u$ and $v$ lie on a plane perpendicular to the direction of the position center. The interferometer measures visibilities between each pair of antennas. For a number of antennas $n$, the number of possible baselines goes as $(n^2 - n)/2$. By increasing the number of antennas, the number of visibility points sampled in the $(u, v)$ plane increases, and thus more information can be inferred about the source intensity distribution on the sky.

However, the observed visibilities $\mathcal{V}_{i,j}$ on a baseline between two antennas $i$ and $j$ are the true visibilities $\mathcal{V}_{i,j}$ corrupted by instrumental and atmospheric effects. The process of correlation and calibration to remove instrumental and atmospheric corruption in EHT data in order to recover $\mathcal{V}_{i,j}$ (also applicable to interferometric data in general) is described in Chapter 5.

For interferometers where the individual antennas are separated by too large distances to correlate the signals in real-time, so far only possible at radio observing wavelengths, the technique of very long baseline interferometry (VLBI) is used. Each antenna is equipped with a very accurate atomic clock (an active hydrogen maser frequency standard, commonly just called ‘maser’) that measures the arrival time of the signal at the antenna. The signal is then time-tagged, downconverted, digitized and recorded to disk, essentially ‘frozen’ until correlation at a later time with the recordings of the other antennas. The technique of VLBI enables imaging at the highest resolutions possible in astronomy, with instruments spanning the entire globe [e.g., Very Long Baseline Array¹ (VLBA), the East Asian VLBI Network², the European VLBI Network³ (EVN) and the EHT; Event Horizon Telescope Collaboration et al., 2019b].

1.4 Interferometric Imaging

Observing at short radio wavelengths is hindered by atmospheric turbulence and water vapor content, which scramble and attenuate the signal received at each antenna. At a wavelength of 1.3 mm, the wavelength at which the emission region around M87* and Sgr A* are expected to be optically thin enough to image their shadows, there are few viable sites on Earth that have dry

¹https://science.nrao.edu/facilities/vlba.
and stable enough atmospheres to fill the \((u, v)\) plane. We thus must use an additional technique to recover information about the intensity distribution on the sky: aperture synthesis via Earth rotation [Ryle & Hewish, 1960; Thompson et al., 2017]. As the Earth rotates, the projection of individual baselines in the \((u, v)\) plane changes, sampling different spatial frequencies of the image as a function of time. Assuming a static intensity distribution during an observation night, we can thus combined visibility measurements over a few hours and fill the \((u, v)\) coverage with the help of the Earth’s rotation (aperture synthesis).

Interferometric imaging is the process of converting the interferometric observables (the visibilities) into an intensity distribution consistent with the observations. The sparser the visibility data in the \((u, v)\) plane the larger the number of images consistent with the observations. Interferometric imaging of short-wavelength VLBI data is an ill-posed problem: there are infinitely many images that are consistent with the observed data.

We can divide interferometric imaging techniques into two categories: inverse and forward modeling techniques. Inverse modeling techniques have been used in the radio community for almost half a century. They consist of first applying an inverse Fourier transform to the visibility data to obtain an intensity distribution called the 'dirty image', which is a convolution of the true intensity pattern on the sky and instrumental effects. The most commonly used inverse modeling algorithm is the CLEAN algorithm, which iteratively reconstructs the image structure using an array of point-like sources and removing instrumental errors [Högbom, 1974; Clark, 1980]. Inverse modeling methods fare very well for interferometric arrays with a well-sampled \((u, v)\) coverage, for which the source intensity pattern would be visible after a simple inverse Fourier transform. For sparse and heterogeneous arrays like the EHT, interferometric imaging may require more flexibility to guide the intensity distribution recovery process and separate the source signal from instrument contributions, flexibility that is not present in inverse modeling algorithms designed for longer-wavelength or fuller-coverage imaging.

In recent years, forward modeling techniques have been developed where the starting point is an intensity distribution that is iteratively constructed, correcting for instrumental errors in phase and amplitude, and Fourier transformed to compare to the visibility data. These techniques allow more flexibility to drive image reconstruction based on assumptions about the underlying source structure and the instrument. The most common framework for forward modeling is regularized maximum likelihood (RML). In RML, source structure is driven by least-squares fits to data products and regularization functions governing particular source properties, such as entropy, smoothness, sparsity, source extent, or temporal continuity [Narayan & Nityananda, 1986; Wiaux et al., 2009a,b; Honma et al., 2014; Chael et al., 2016, 2018b; Bouman et al., 2016, 2018; Akiyama et al., 2017a,b; Johnson et al., 2017; Kuramochi et al., 2018, Chapter 3]. Many of the RML and image validation developments and were made specifically to surmount the challenges of imaging EHT data, see Event Horizon Telescope Collaboration et al. [2019d] and Chapter 6 for further detail.

1.5 A Tale Of Two Black Holes

Two supermassive black holes have angular sizes on the sky sufficiently large to be resolvable with an Earth-sized synthesized telescope: the supermassive black hole at the center of our Milky Way,
1.5 A Tale Of Two Black Holes

Sagittarius A* (Sgr A*), and the supermassive black hole at the center of the M87 giant elliptical galaxy, M87* [Doeleman et al., 2009a; Goddi et al., 2017]. In the mid-2000s, a collaborative effort begins to combine telescopes at high and dry sites observing in the millimeter radio regime, in order to reach the resolution necessary to detect structure in those two sources on event horizon scales.

1.5.1 Sagittarius A*

The radio source Sgr A* was detected for the first time by Balick & Brown [1974] in Sagittarius A, in our Milky Way galactic center. The position of the radio source Sgr A* [Menten et al., 1997; Reid & Brunthaler, 2004; Reid, 2009] coincides with the center of gravity around which nearby S-stars, observed in the near-infrared, orbit in highly elliptical orbits [see Figure 1.7; Eckart & Genzel, 1997; Ghez et al., 1998]. The gravitational effects on the stellar orbits inferred that the central object, not detectable in the near-infrared, must be extremely compact and extremely massive. With the mass and distance measurements inferred from the S-stars with the Gravity experiment and the Keck telescopes, the radio source is now associated with a supermassive black hole candidate of a mass \( M \sim 4 \times 10^6 M_\odot \) at a distance \( D \sim 8 \) kpc [Ghez et al., 2008; Gillessen et al., 2009, 2017; Gravity Collaboration et al., 2018a, 2019; Do et al., 2019]. The angular size of the shadow of the black hole Sgr A* is thus estimated to be \( \sim 50 \) μas.

Observations of Sgr A* in the radio and infrared regimes showed that its spectral energy distribution rises with frequency in the radio due to synchrotron emission, with a turnover at \( \sim 1 \) THz (submillimeter wavelengths), where the accretion flow becomes optically thin [see Figure 1.8; Falcke et al., 1998; Bower et al., 2015a, 2019]. Its bolometric luminosity was measured
to be $\sim 5 \times 10^{35}\text{erg s}^{-1}$, or $10^{-9}L_{\text{Edd}}$ [Bower et al., 2019]. The radio spectrum of Sgr A* however seems to exhibit some ambiguity: the dominant mechanism to produce the observed synchrotron emission could either be a compact relativistic jet [Falcke & Markoff, 2000] or a RIAF [Narayan et al., 1995; Özel et al., 2000; Yuan et al., 2003]. There is also evidence of frequency-dependent time lags in radio light-curves of Sgr A* that may be indicative of an outflow [Yusef-Zadeh et al., 2006, 2008; Brinkerink et al., 2015, Brinkerink et al., 2021 in prep.] but no outflow or jet has ever been imaged in the radio (unlike in other LLAGN like M87*).

At long radio wavelengths, imaging observations of Sgr A* are dominated by effects of interstellar scattering by free electrons in the ionized interstellar medium (ISM) between us and the Galactic Center (see Chapters 1.6, 3 and 5 for further discussion on scattering). Due to scattering, the image of Sgr A* is blurred, with a $\lambda^2$ apparent size dependence and an anisotropic broadening predominantly along the east-west axis [Davies et al., 1976; van Langevelde et al., 1992; Alberdi et al., 1993; Marcaide et al., 1999; Bower et al., 2004, 2006, 2014; Shen et al., 2005; Lu et al., 2011a; Johnson et al., 2018]. Images of the intrinsic structure in Sgr A* have thus been very challenging for decades, and led to the development of imaging instruments at shorter wavelengths.

The first VLBI Sgr A* detection at 1.4 mm by Krichbaum et al. [1998] shows a compact source on a single baseline, with potential for expansion. In 2007, Doeleman et al. [2008] conduct the first 1.3 mm VLBI observations of Sgr A* to detect the source between multiple pairs of telescopes (or ‘baselines’, see Chapter 1.3), between Hawaii, Arizona, and California. They detect structure on event horizon scales, and measured a source size comparable to the expected size of the shadow as predicted by Einstein’s theory of GR ($\sim 50\mu\text{as}$). With observations from 2009, Fish et al. [2011] are the first to detect the source on three baselines simultaneously. With three baselines, ‘closure phases’, the sum of phases on a baseline triangle, are constructed and deviations from zero indicate the presence of asymmetric structure. They also show that the source exhibits variability within

Figure 1.8: Spectral energy distribution for Sgr A* in the radio and infrared regimes. The spectral energy density in the radio rises with frequency, with a turnover in the sub-millimeter wavelength regime. Figure from Melia & Falcke [2001].
a single observing night. Johnson et al. [2015], using 2013 observations with US stations, are the first to measure polarized structure in Sgr A* on horizon scales, and show that the structure indicates fairly ordered magnetic fields. The following year, Fish et al. [2016] combine closure phases across multiple years of observations and show that the asymmetric structure in Sgr A* is persistent over years, but uncertainty remains around its origin: it can be either intrinsic structure or contamination from interstellar scattering (see Chapter 1.6). In 2018, Lu et al. [2018] published results from 2013 observations not only with the US stations but with an added station in Chile. The added baselines to Chile brought an unprecedented view into the Sgr A* source structure: for the first time, observations on horizon scales are no longer consistent with a Gaussian intensity distribution on the sky, but match that of a ring annulus.

However, while scattering effects are minimal in the millimeter/submillimeter regime, the variability of Sgr A* proves to be the greatest challenge for EHT observations. Due to its mass, Sgr A* varies on timescales of minutes. Over the course of an observation night, the image of Sgr A* can undergo drastic structural changes that may not be trackable with a sparse interferometric array, or imageable with standard imaging techniques (see Chapter 3). Furthermore, in the case of images of M87* with the EHT [Event Horizon Telescope Collaboration et al., 2019a], the optically thin emission at 1.3 mm originates from very close to the black hole. Extended emission is weaker than the peak intensity of the near-horizon emission by a factor of \( > 10 \), and is difficult to constrain with our current instrument. We expect a similar imaging dynamic range for Sgr A*. With the current EHT array, it is thus expected that the image will be dominated by the gravitational lensing of the gas close to the black hole, and signatures of an outflow or jet would be very difficult to constrain. The ideal regime to constrain the dominating radio emission mechanism is thus further out in the emission region, where the emission is optically thicker and easier to detect and recover via direct imaging. The optimal observing wavelength to carry out such observations is at 3.5 mm, where signatures from outflows or jets can be imaged, and interstellar scattering effects are of the same order as the intrinsic source extent, and thus separable via scattering modeling techniques [Johnson, 2016; Johnson et al., 2018]. History and studies of the image of Sgr A* at 3.5 mm are the subject of Chapters 2 and 4.

1.5.2 M87*

The relativistic jet in M87 was first observed in the optical band by Heber Curtis in 1918 with the Lick Observatory, who describes it as a “peculiar thin bright stream of matter originating from a compact source” [Curtis, 1918]. This jet extends out of the M87 galaxy across a distance of \( \sim 65 \) kpc and emits radiation in radio, optical, and X-ray bands [Owen et al., 1989, 2000; Sparks et al., 1996; Perlman et al., 1999; Marshall et al., 2002; de Gasperin et al., 2012]. Measured core shifts from multi-wavelength observations of the jet across the radio band indicate that the base of the jet should be attainable at 1.3 mm [see Figure 1.9; e.g., Kovalev et al., 2007; Ly et al., 2007; Asada et al., 2014; Hada et al., 2017; Walker et al., 2018; Kim et al., 2018]. The central engine that powers this jet is believed to be the supermassive black hole at the center of the M87 galaxy, at a distance of \( \sim 17 \) Mpc [Blakeslee et al., 2009; Bird et al., 2010; Cantiello et al., 2018]. The mass of the black hole is indirectly measured to be \( \sim 3 \times 10^9 M_\odot \) via gas dynamics [Walsh et al., 2013] and \( \sim 6 \times 10^9 M_\odot \) via stellar dynamics [Gebhardt et al., 2011]. This range in the black
Chapter 1: Introduction

Figure 1.9: A multi-wavelength view of M87* and its radio jet. The resolution of the EHT gives us an unprecedented view of the environment in the vicinity of a supermassive black hole. Individual images are from Event Horizon Telescope Collaboration et al. [2019a], Kim et al. [2018], Walker et al. [2018], NRAO/VLA.

Figure 1.10: First images of the shadow of the supermassive black hole M87* observed at 1.3mm with the Event Horizon Telescope on four independent days during the 2017 observing campaign. Figure from Event Horizon Telescope Collaboration et al. [2019d].

hole mass adds uncertainty as to whether a shadow is resolvable with an Earth-sized array: the lower mass end would give a black hole shadow size of order of the instrument resolution in the millimeter radio ($\psi \sim 20 \mu$as), with poor prospects for imaging.

The first 1.3mm VLBI detection of M87* in 2009, by Doeleman et al. [2012], reveal source structure on horizon scales ($\sim 40 \mu$as). Uncertainty in the black hole mass, due to competing measurements from stellar dynamics [Gebhardt et al., 2011] and gas dynamics [Walsh et al., 2013], keep expectations low for the possibility of resolving the M87* shadow with an Earth-sized interferometer at 1.3mm observing wavelength. Further observations with the early EHT by Akiyama et al. [2015] confirmed the size of the source and measured the first closure phases on M87*: these were consistent with zero, the signature of a symmetric source structure. The expansion of the EHT and the addition of the Atacama Millimeter/submillimeter Array (ALMA), the world’s most advanced radio-astronomy instrument, leads to an imaging-capable array in 2017. In April 2017, a fully fledged EHT carries out observations of M87* and Sgr A* (as well as other well known astrophysical black holes) with eight single-dish telescopes and arrays at six different locations across the globe. After two years of processing and scientific analysis, the 2017 observations led to the first images of a black hole: the EHT reveals to the world the shadow of the supermassive black hole M87* [see Figure 1.10; Event Horizon Telescope Collaboration et al., 2019a,b,c,d,e,f]. The shape of the shadow deviates from circularity by < 10% and is consistent with the prediction from Einstein’s theory of GR. The mass of the black hole, inferred from the ring size measurement across multiple techniques, was found to be $(6.5 \pm 0.7) \times 10^9 M_\odot$, consistent with the stellar dynamics mass measurement from Gebhardt et al. [2011]. Further analysis of
Figure 1.11: Final polarimetric images of M87* observed at 1.3 mm with the EHT on four independent days during the 2017 observing campaign, shown with two visualization schemes. Top: Total intensity is shown in grayscale, the tick color represents polarization fraction, and tick direction represents the electric vector position angle (EVPA). The tick length is scaled as the square of the polarized intensity. Bottom: Polarization “field lines” tracing EVPA patterns plotted atop an underlying total intensity image, treating the linear polarization as a vector field. The length and opacity of the streamlines are scaled as the square of the polarized intensity. Figure from Event Horizon Telescope Collaboration et al. [2021a].
full-polarization observations led to the first polarized images of the M87* black hole [Event Horizon Telescope Collaboration et al., 2021a,b]. The polarized emission reaches a peak of 15\%, suggesting Faraday and beam depolarization, and the EVPAs lie in a spiral pattern, from which a spiral magnetic field configuration with a poloidal component is inferred. The magnetic field is moderately strong for a black hole of the size of M87*, its strength measured to be 4 – 30 Gauss. The calibration, imaging, and validation processes that led to the total-intensity and polarimetric images of M87* with the EHT are the subject of Chapters 5 and 6.

1.6 Interstellar Scattering

The effects of interstellar scattering on radio observations of Sgr A* have long hindered the recovery of intrinsic source structure, making it challenging to determine the primary origin of the radio emission (disk or jet) in the source. Currently, tests of GR via black hole imaging are limited by astrophysical uncertainties of the surrounding emission region and the accuracy of the black hole mass measurement [Event Horizon Telescope Collaboration et al., 2019e,f]. While the precision of the mass and distance measurements of Sgr A* make it the ideal candidate to test theories of gravity, interstellar scattering toward the Galactic Center is a barrier to understanding fundamental properties of the source that would enable such tests with the EHT.

Density inhomogeneities in the ionized ISM scatter radio waves, causing temporal and angular broadening and scintillation of observed sources [e.g., Rickett, 1990]. Across a single path length $\delta z$, an electron density fluctuation $\delta n_e$ would give a phase change $\delta \phi = -r_e \lambda \times \delta z \times \delta n_e$, where $r_e$ is the classical electron radius [see e.g., Thompson et al., 2017]. Along many lines of sight, these density variations can be characterized by a stochastic phase $\phi(r)$, where $r$ is a 2D vector transverse to the line of sight, changing across a thin scattering screen. These fluctuations are characterized by a dimensionless power spectrum $Q(q) = |q|^{-\beta}$, where $q$ is the wavenumber, with an unbroken power law between an outer (or injection) scale $r_{\text{out}}$ and an inner (or dissipation) scale $r_{\text{in}}$, suggesting a turbulent cascade [e.g., Tatarskii, 1971; Armstrong et al., 1995; Lambert & Rickett, 1999]. For a Kolmogorov spectrum of density fluctuations, we have $\beta = 11/3$ [Goldreich & Sridhar, 1995].

We quantify the effects of these phase variations on interferometric visibilities with the phase structure function,

$$D_\phi(r) = 2[C(0) - C(r)] = \langle [\phi(r' + r) - \phi(r')]^2 \rangle \propto \lambda^2,$$

where $C(r)$, the phase correlation function, is related to the power spectrum of phase fluctuations $Q(q)$ by [e.g., Psaltis et al., 2018]

$$C(r) = \lambda^2 \int Q(q) e^{iqr} dq.$$

The structure function takes the shape of an unbroken power law $D_\phi(r) \propto |r|^\alpha$ over scales between $r_{\text{out}}$ and $r_{\text{in}}$, where $\alpha = \beta + 2$. For displacements on these scales, the angular broadening $\theta_{\text{scatt}} \propto \lambda^{1+\frac{2}{\alpha}}$ causes non-Gaussian blurring. At long observing wavelengths, the image is only sensitive to small phase displacements (smaller than $r_{\text{in}}$). The phase structure function is then quadratic (we have $D_\phi(r) \propto r^2 \lambda^2$) and the diffractive kernel becomes Gaussian [Tatarskii, 1971].
The parametrization of the thin scattering screen used in this work is presented in detail in Psaltis et al. [2018], see the schematic in Figure 1.12. The geometry of the scattering screen (the source-screen and observer-screen distances) is based on temporal observations of the Galactic Center magnetar, 2.4" from Sgr A* on the sky in combination with the angular broadening observed across the radio waveband. By comparing the scattering properties of Sgr A* and the nearby magnetar, the position of the common scattering screen with respect to the observer and the Galactic Center is inferred [Bower et al., 2015b]. A model for the underlying magnetic field wander of the screen is assumed, which governs the specific properties of the scattering anisotropy observed [Goldreich & Sridhar, 1995]. In this work we adopt the ‘dipole’ model, see Psaltis et al. [2018] for a discussion of various models.

Phase fluctuations on the diffractive scale induce scatter-broadening. The diffractive scale $r_{\text{diff}}$ is defined as the transverse length on the scattering screen over which the phase structure function becomes equal to unity: $D_{\phi}(r_{\text{diff}}) \equiv 1$. The diffractive scale is related to the angular broadening as

$$r_{\text{diff}} \sim \frac{\lambda}{(1 + M) \theta_{\text{scatt}}}.$$  \hspace{1cm} (1.16)

where $M = D/R$ is the magnification factor for a screen at a distance $D$ from the observer and at a distance $R$ from the source, see Figure 1.12. The scattering kernel transitions from Gaussian to non-Gaussian for $b \sim (1 + M)r_{\text{in}}$, where $b$ is a vector baseline of an interferometer. For baselines $b \sim r_{\text{diff}}$, sensitive to scatter-broadening, the kernel visibility function falls off. When $r_{\text{diff}} < r_{\text{in}}$, the scatter-broadening is thus Gaussian [Johnson & Gwinn, 2015]. For some intuition, observations of Sgr A* at long centimeter wavelengths exhibit a $\lambda^2$ scattering behavior and a
Figure 1.13: Measured scattered image major axis size (FWHM) as a function of wavelength. At centimeter wavelengths, the image size exhibits a $\lambda^2$ behavior, but deviates in the millimeter regime. This deviation can be modeled by a wavelength dependence of both the scatter-broadening kernel size and the intrinsic source size. Figure adapted from Johnson et al. [2018].

Gaussian scattering kernel, for which $r_{\text{diff}} < r_{\text{in}}$ and $\theta_{\text{source}} < \theta_{\text{scatt}}$ [e.g., Bower et al., 2004, 2006, see Figure 1.13]. At millimeter wavelengths, particularly for the EHT, we expect $r_{\text{diff}} \gtrsim r_{\text{in}}$ and $\theta_{\text{source}} \gtrsim \theta_{\text{scatt}}$, thus scattering modeling allowing for deviations from $\lambda^2$ is needed [see Figure 1.14; Psaltis et al., 2018; Johnson et al., 2018]. For Sgr A*, the expected dissipation scale ($r_{\text{in}}$) in the ISM is in the range of $10^2 - 10^3$ km [e.g., Spangler & Gwinn, 1990], so the transition to non-Gaussian and non-$\lambda^2$ scattering is expected to occur at wavelengths of a few millimeters [Johnson et al., 2018].

While fluctuations on the diffractive scale produce angular broadening, phase fluctuations comparable to the refractive scale $r_{\text{ref}}$ cause refractive scintillation. The refractive scale $r_{\text{ref}}$ corresponds to the projected size of angular broadening on the scattering screen,

$$\text{r}_{\text{ref}} \equiv \frac{r_F^2}{r_{\text{diff}}} \sim \theta_{\text{scatt}} D, \quad (1.17)$$

where the purely geometric Fresnel scale is given by

$$r_F = \sqrt{\frac{\lambda DR}{2\pi(D+R)}}. \quad (1.18)$$

These modes are much larger than $r_{\text{diff}}$, so while two scattering prescriptions may have identical scatter-broadening, they may differ in their scattering substructure induced by refractive modes, and could potentially have huge consequences on tests of GR with EHT images of Sgr A* [e.g., Zhu et al., 2019]. Refractive scattering, first proposed by Narayan & Goodman [1989] and Goodman & Narayan [1989], and first detected in Sgr A* by Gwinn et al. [2014], not only introduces image wander and flux density modulation but also introduces stochastic substructure in the
image. This image substructure translates to added time and baseline-dependent complex ‘noise’ to interferometric visibilities [Johnson & Gwinn, 2015; Johnson & Narayan, 2016]. A detailed derivation of the added refractive noise for VLBI observations is presented in Johnson et al. [2018].

In the regime of strong scattering – where \( r_{\text{ref}} > r_F > r_{\text{diff}} \), the case for Sgr A* below THz frequencies – Narayan & Goodman [1989] and Goodman & Narayan [1989] present the effects of scattering on visibilities in three averaging regimes: the ‘snapshot-image’ regime, the ‘average-image’ regime, and the ‘ensemble-average’ regime. The timescales associated with each averaging regime depend on the integration time and fractional bandwidth of the instrument. In the snapshot image regime, stochastic variations in frequency and time will affect the visibility measurements due to diffractive scintillation (this regime is not applicable for extended sources). In the average-image regime, short-timescale diffractive fluctuations are averaged out, but visibilities still fluctuate in frequency and time due to refractive scintillation (this regime corresponds to a single refractive scattering realization, and is applicable on timescales of minutes to an entire observing night). In the ensemble-average regime, integrating over long timescales and wide bandwidths will simply result in the intrinsic (unscattered) image convolved with a diffractive kernel (this timescale corresponds to the average of many refractive scattering realizations, over many observing nights). For a point source, the Fourier visibilities \( \mathcal{V}(\boldsymbol{u}) \) observed by an interferometer in the ensemble-average regime are given by

\[
\mathcal{V}(\boldsymbol{u}) = \exp \left[ -\frac{1}{2} D_\phi \left( \frac{\lambda \boldsymbol{u}}{1 + M} \right) \right],
\]

(1.19)

where \( \boldsymbol{u} = (u, v) \). Observations of Sgr A* presented in this thesis are in the average-image regime,
Chapter 1: Introduction

\[ \alpha = 1.0 \]
\[ r_{\text{in}} = 10 \text{ km} \]

\[ \alpha = 1.45 \]
\[ r_{\text{in}} = 10 \text{ km} \]

\[ \alpha = 1.9 \]
\[ r_{\text{in}} = 10 \text{ km} \]

\[ \alpha = 1.0 \]
\[ r_{\text{in}} = 10^3 \text{ km} \]

\[ \alpha = 1.45 \]
\[ r_{\text{in}} = 10^3 \text{ km} \]

\[ \alpha = 1.9 \]
\[ r_{\text{in}} = 10^3 \text{ km} \]

\[ \alpha = 1.0 \]
\[ r_{\text{in}} = 10^5 \text{ km} \]

\[ \alpha = 1.45 \]
\[ r_{\text{in}} = 10^5 \text{ km} \]

\[ \alpha = 1.9 \]
\[ r_{\text{in}} = 10^5 \text{ km} \]

Figure 1.15: Modeled effects of the interstellar scattering expected for Sgr A* on a simulated uniform ring with an outer diameter of 55 μas and an inner diameter of 45 μas at λ = 1.3 mm with varying power-law index α and inner scale of turbulence r_{\text{in}}. The solid circle shows the central ring diameter of 50 μas. Each row has constant α and varying r_{\text{in}}, while each column has varying α and constant r_{\text{in}}. The degeneracy between the two parameters is shown. The scattering parametrization used is that of Psaltis et al. [2018] in the Johnson [2016] modeling framework. Based on the work presented in this thesis, the center panel is a good approximation for our current understanding of the scattering screen.
where diffractive fluctuations are suppressed but refractive fluctuations, along with angular broadening, affect our interferometric measurements.

The properties of the scattering screen toward Sgr A* thus essentially come down to constraining six quantities: the scatter-broadening kernel major and minor axis full-widths at half maximum $\theta_{\text{maj,0}}$ and $\theta_{\text{min,0}}$ and position angle $\phi_{\text{PA}}$, the power law index $\alpha$, the inner scale $r_{\text{in}}$ and the outer scale $r_{\text{out}}$. For Sgr A*, the outer scale $r_{\text{out}}$ is much larger than any scale probed with observations, and is thus negligible. The scattering kernel parameters are well constrained by Johnson et al. [2018] using multi-wavelength observations of Sgr A* across the radio waveband, from 20 cm to 1.3 mm (though millimeter-wave observations were limited). Estimates for $\alpha$ and $r_{\text{in}}$ were derived, although due to the lack of refractive scattering detections in the millimeter regime (from limitations in resolution and/or instrument sensitivity), the transition between centimeter-wave scattering properties and millimeter-wave properties is not trivial. Due to the uncertainty in $\alpha$ and $r_{\text{in}}$, and the degeneracy in angular broadening and refractive scattering properties induced by the two parameters, expectations for images with the EHT in the millimeter regime could vary wildly and hinder tests of GR, see Figure 1.15. Observations in the millimeter regime, with increased sensitivity and resolution, can help break degeneracies in the scattering parameters, and are the subject of Chapters 2 and 4.

1.7 This Thesis

In this thesis, I present five chapters related to observations of two EHT primary targets with high-frequency VLBI: Sagittarius A* and M87*. The first part describes analysis and scientific developments in 86 GHz observations of Sagittarius A* with the Global Millimeter VLBI Array (GMVA), while the second part describes technique developments for 230 GHz observations of M87 with the Event Horizon Telescope.

In Chapter 2, the first observations of Sagittarius A* at 86 GHz with the Atacama Millimeter-/submillimeter Array (ALMA) as part of a VLBI array are presented. The observations enabled the imaging of the intrinsic structure in Sgr A*, yielding a source structure that deviates modestly from circular symmetry. Comparisons with a set of theoretical simulations showed that jet-dominated models for its radio emission inclined beyond $20^\circ$ of the line of sight are too elongated to match the results.

In Chapter 3, an imaging technique using the second moment properties of the visibility function for compact sources is laid out. On short baselines, a compact source will appear Gaussian. For sparse arrays lacking short baselines, or arrays in which short baselines have calibration errors, the source extent (or second moment) constraint aids to contain flux density in the imaging process and converge quicker to an image reconstructed with high fidelity. This technique helped offset calibration errors in the imaging of the 86 GHz GMVA+ALMA data set without data manipulation, and proves to be an effective technique to constrain intensity distribution in movie reconstructions with the EHT.

In Chapter 4, an analysis of 2017 and 2018 86 GHz observations of Sagittarius A* is presented. The combined properties of the three data sets showed that the image of Sgr A* is persistently non-Gaussian on the sky, and its extent and asymmetry remain consistent with disk-dominated emission models and low-inclination jet dominated-models. An analysis of the scattering proper-
ties of the data sets showed that scattering constraints inferred from 86 GHz results are consistent with those from lower frequencies, putting confidence in a single scattering model to describe the ISM behavior across the radio band.

In Chapter 5, the data calibration and processing pathway for 2017 EHT observations at 230 GHz is presented. The high bandwidth and heterogeneity of the EHT data brought many challenges to standard VLBI data reduction procedures, and required the development of new pipelines and tools to address them. We utilized three phase calibration pipelines: the Haystack Observatory Processing System [HOPS; Whitney et al., 2004; Blackburn et al., 2019]; the Common Astronomy Software Package (CASA) pipeline rPICARD [McMullin et al., 2007; Janssen et al., 2019b]; and the Astronomical Image Processing System (AIPS) pipeline developed especially for handling of EHT data with standard low-frequency VLBI tools, whose development I co-led [Greisen, 2003; Event Horizon Telescope Collaboration et al., 2019c]. We further developed a framework in which the output of all three phase calibration pipelines could undergo polarimetric and amplitude calibration [Blackburn et al., 2019]. Validation procedures were also developed for data quality assurance across pipelines, bands, and data products. This pathway was used for the scientific M87 results released by the EHT collaboration in April 2019.

In Chapter 6, total intensity and imaging validation procedures are presented. The imaging of EHT data required the development of new imaging techniques and workflows, the use of parameter surveys for systematic imaging assessment, and validation of derived quantities across methods, bands, and observed sources. These procedures were crucial in the path toward the first image of a black hole and the first polarized image of a black hole [Event Horizon Telescope Collaboration et al., 2019d, 2021a]. In the total-intensity imaging efforts, I co-led one of four imaging teams at the blind imaging stage, contributed to the organization of imaging rounds and parameter surveys, and led the effort on data and instrument properties derived from the images. In the polarimetric imaging efforts, I led the organization of the entire polarimetric effort into software teams, ran synthetic and real data challenges, prepared comparison frameworks and polarimetric image visualizations, and co-led one of three software teams and parameter surveys. For both total intensity and polarimetric imaging publications, I was a leading member of the writing team.
The Size, Shape, and Scattering of Sagittarius A* at 86 GHz: First VLBI with ALMA


Abstract

The Galactic Center supermassive black hole Sagittarius A* (Sgr A*) is one of the most promising targets to study the dynamics of black hole accretion and outflow via direct imaging with very long baseline interferometry (VLBI). At 3.5 mm (86 GHz), the emission from Sgr A* is resolvable with the Global Millimeter VLBI Array (GMVA). We present the first observations of Sgr A* with the phased Atacama Large Millimeter/submillimeter Array (ALMA) joining the GMVA. Our observations achieve an angular resolution of $\sim 87\mu$as, improving upon previous experiments by a factor of two. We reconstruct a first image of the unscattered source structure of Sgr A* at 3.5 mm, mitigating effects of interstellar scattering. The unscattered source has a major axis size of $120 \pm 34\mu$as ($12 \pm 3.4$ Schwarzschild radii), and a symmetrical morphology (axial ratio of $1.2^{+0.3}_{-0.2}$), which is further supported by closure phases consistent with zero within $3\sigma$. We show that multiple disk-dominated models of Sgr A* match our observational constraints, while the two jet-dominated models considered are constrained to small viewing angles. Our long-baseline detections to ALMA also provide new constraints on the scattering of Sgr A*, and we show that refractive scattering effects are likely
to be weak for images of Sgr A* at 1.3 mm with the Event Horizon Telescope. Our results provide the most stringent constraints to date for the intrinsic morphology and refractive scattering of Sgr A*, demonstrating the exceptional contribution of ALMA to millimeter VLBI.

2.1 Introduction

Supermassive black holes (SMBHs) play a crucial role in shaping our Universe: they evolve symbiotically with their host galaxies and are the cause of extreme environmental changes via accretion, outflows, jets and mergers [e.g., Ferrarese & Merritt, 2000; Gebhardt et al., 2000]. They are believed to be the origin of the most energetically efficient and powerful processes in the Universe, and yet we are far from fully grasping how these processes are launched and maintained [e.g., Boccardi et al., 2017; Padovani et al., 2017]. Several theories have been put forward to explain accretion and jet launching mechanisms of SMBHs, but observational evidence to discriminate among theoretical models remains scarce [e.g., Yuan & Narayan, 2014; Fragile, 2014].

Sagittarius A* (Sgr A*) is the radio source associated with the closest known SMBH, with a mass $M \sim 4.1 \times 10^6 M_\odot$, located at the center of our Milky Way, at a distance $D \sim 8.1$ kpc [Ghez et al., 2008; Reid, 2009; Gillessen et al., 2009; Gravity Collaboration et al., 2018a]. The angular size of the Schwarzschild radius for Sgr A* is thus estimated to be $R_{\text{Sch}} = 2GM/c^2 \sim 10\mu$as. Due to its proximity, Sgr A* subtends the largest angle on the sky among all known SMBHs, and is thus the ideal laboratory to study accretion and outflow physics [Goddi et al., 2017].

Theoretical models of the dominating component of the radio emission in Sgr A* fall into two broad classes: a relativistic compact jet model or a radiatively inefficient accretion flow [Narayan et al., 1995; Falcke & Markoff, 2000; Özel et al., 2000; Yuan et al., 2003]. However, the southern declination and strong interstellar scattering of Sgr A* (see more details in Section 2.2.2) lead to uncertainty in its intrinsic radio structure, despite decades of centimeter wavelength very long baseline interferometry (VLBI) observations [e.g., Alberdi et al., 1993; Marcaide et al., 1999; Bower et al., 2004; Shen et al., 2005; Lu et al., 2011b; Bower et al., 2014]. Consequently, these observations have so far been unable to decisively constrain the dominating emission model for Sgr A* to either of those two classes. Additional lines of evidence provide support for both models. For instance, frequency-dependent time lags in light-curves of Sgr A* suggest expanding outflows during flares [e.g., Yusef-Zadeh et al., 2006, 2008; Brinkerink et al., 2015]. VLBI observations at 7 mm have found evidence for significant intrinsic anisotropy in some epochs [Bower et al., 2014], although the anisotropy is not universally seen for other instruments and epochs [e.g., Zhao et al., 2017], so the anisotropy may be episodic or may be due to limitations in the scattering mitigation or model fitting procedure.

VLBI in the mm-regime can reach the smallest spatial scales in Sgr A*, enabling detection and imaging of the intrinsic structure. At a wavelength of 1.3 mm, observations with the Event Horizon Telescope (EHT) have shown that the radio emission occurs on scales comparable to the event horizon [Doeleman et al., 2008; Fish et al., 2011; Johnson et al., 2015; Fish et al., 2016; Lu et al., 2018]. On these scales, general relativistic effects such as the “shadow” cast by the black hole are expected to determine the source morphology [Falcke et al., 2000], limiting the view of the
innermost accretion flow. At longer wavelengths, scatter-broadening by the interstellar medium (ISM) strongly hinders any attempt to probe intrinsic structure. Observations at 3.5 mm, where accretion flow kinematics may give rise to an outflow or compact jet and where scatter-broadening becomes subdominant to intrinsic structure, can distinguish between the two classes of models via detailed comparisons of observations and simulations, and help understand the fundamental nature of the radio emission from Sgr A*.

The first 3.5 mm VLBI detection of Sgr A*, by Rogers et al. [1994], gave an initial estimate of the scattered source size using a circular Gaussian fit. Krichbaum et al. [1998] used three stations to measure the first closure phases (consistent with zero) at 3.5 mm on a small triangle. Closure phases are a robust observable, since the closed sum of phases in a triangle removes any station-based instrumental effect. A zero value indicates symmetry in the spatial scales probed by the three baselines involved in the closure measurement, a non-zero value implies asymmetry [e.g., Rauch et al., 2016; Thompson et al., 2017]. Subsequent observations, with improved sensitivity and baseline coverage, used closure amplitudes for elliptical Gaussian model-fitting, but the minor axis of the scattered source, along the north-south direction, remained difficult to constrain because of predominantly east-west array configurations [Doeleman et al., 2001; Shen et al., 2005; Bower et al., 2006; Lu et al., 2011b].

The addition of the Large Millimeter Telescope Alfonso Serrano (LMT) and the Robert C. Byrd Green Bank Telescope (GBT) enabled more precise estimates of the intrinsic size and shape of Sgr A* and revealed non-zero closure phases, indicating either intrinsic source asymmetry or substructure from interstellar scattering [Ortiz-León et al., 2016; Brinkerink et al., 2016, hereafter O16, B16]. Further analysis by Brinkerink et al. [2019, hereafter B18] found a slight excess of flux density (≈1% of total flux density) east of the phase center, giving clear deviation from the purely Gaussian geometry that was assumed in model-fitting. Thus, these improved observations support moving beyond simple Gaussian model-fitting to test more complex source models. Imaging is a natural next step, as it does not assume a particular morphological model.

The development of phased-array capability at the Atacama Large Millimeter/submillimeter Array (ALMA) gives unprecedented sensitivity at 3.5 mm [Doeleman, 2010; Fish et al., 2013; Matthews et al., 2018]. In addition to its sensitivity, the geographical location of ALMA provides long north-south baselines to Northern hemisphere sites, probing regions where scattering is subdominant to intrinsic structure. In this paper, we present the first VLBI observations of Sgr A* with phased ALMA joining twelve stations of the Global Millimeter VLBI Array (GMVA). These observations improve north-south resolution by more than a factor of three compared to previous 3.5 mm experiments, and they allow us to reconstruct the first unscattered image of Sgr A* at 3.5 mm.

The organization of the paper is as follows. In Section 2.2, we give an overview of the relevant background for models of the intrinsic structure and scattering of Sgr A*. After summarizing the observations and data reduction (Section 2.3) and the imaging (Section 2.4), we present our GMVA+ALMA image and discuss data- and image-derived properties of the intrinsic source in the context of previous 3.5 mm experiments in Section 2.5. In Section 2.6, we discuss our new constraints on theoretical models for Sgr A* and its scattering. We summarize our results in Section 2.7.
2.2 Background

2.2.1 Theoretical models for Sgr A* emission

Sgr A* is a bright radio source, with a spectrum that rises with frequency until it peaks near 1 mm [e.g., Falcke et al., 1998; Bower et al., 2015]. The long-standing debate on whether the radio/mm emission from Sgr A* is produced by a radiatively inefficient accretion disk or by a relativistic, compact jet present near the black hole [e.g., Narayan et al., 1995; Markoff et al., 2007; Mościbrodzka et al., 2014; Ressler et al., 2015; Connors et al., 2017; Davelaar et al., 2018; Chael et al., 2018a, and references therein] has not been resolved.

Radiative models of Sgr A* based on three-dimensional general relativistic magnetohydrodynamic (GRMHD) simulations of Kerr black hole accretion naturally combine the disk and jet scenarios. Electrons and ions are not in thermal equilibrium in the hot, diffuse Sgr A* accretion flow, therefore simulations with the same gas dynamics (determined by the ions) can have quite different appearances at 3.5 mm depending on electron thermodynamics assumptions. In particular, both the disk and jet emission dominated models can be realized within a single simulation by adopting a specific distribution for electron heating/acceleration in magnetized plasma in post-processing [e.g., Mościbrodzka & Falcke, 2013]. Alternatively, electron-ion thermodynamics with a specified prescription for the particle heating from dissipation can be incorporated self-consistently with the other variables in a single simulation. In this framework, Ressler et al. [2017] and Chael et al. [2018a] have shown that both jet- and disk-dominated images can be produced at 3.5 mm, depending on the underlying physical model for electron heating evolved in the simulation.

These models are mainly used to predict 1.3 mm EHT observations [e.g., Chan et al., 2015]. At 1.3 mm we expect the emission to originate near the event horizon where effects such as gravitational lensing and relativistic Doppler boosting distort any emission into a ring, crescent or a spot-like shape, making any distinction between dominating emission models difficult. At 3.5 mm, we can potentially constrain the geometry and electron micro-physics of the GRMHD simulations by modeling emission maps in which the physics of accretion rather than relativistic effects shapes the source geometry.

2.2.2 Interstellar Scattering of Sgr A*

The index of refraction of a plasma depends on density, so density inhomogeneities in the ionized ISM lead to multi-path propagation of radio waves. The scattering is chromatic, with scattering angles proportional to the squared wavelength of a propagating wave. Because the scattering arises from density irregularities, scattering properties are stochastic by nature; their statistical properties depend on the power spectrum $Q(q)$ of density variations, where $q$ denotes a wavevector. Along many lines of sight, the scattering is well characterized using a simplified description in which the scattering material is confined within a single thin screen along the line of sight. For background and reviews on interstellar scattering, see Rickett [1990], Narayan [1992], or Thompson et al. [2017].

The line of sight to Sgr A* is particularly heavily scattered, as is evidenced by an image with a Gaussian shape and a size that is proportional to wavelength squared for wavelengths $\lambda \gtrsim 1$ cm
[Davies et al., 1976; van Langevelde et al., 1992; Bower et al., 2004; Shen et al., 2005; Bower et al., 2006; Johnson et al., 2018]. In addition, the scattering of Sgr A* is anisotropic, with stronger angular broadening along the east-west axis than along the north-south axis [Frail et al., 1994]. The angular broadening has a full width at half maximum (FWHM) of $(1.380 \pm 0.013) \lambda_{\text{cm}}^2$ mas along the major axis and $(0.703 \pm 0.013) \lambda_{\text{cm}}^2$ mas along the minor axis, with the major axis at a position angle $81.9^\circ \pm 0.2^\circ$ east of north [Johnson et al., 2018, hereafter J18]. For comparison, the intrinsic source has an angular size of $\sim 0.4 \lambda_{\text{cm}}$ mas [J18], so the ratio of intrinsic size to scatter broadening is $\sim 0.3 / \lambda_{\text{cm}}$ along the major axis and $\sim 0.6 / \lambda_{\text{cm}}$ along the minor axis. Consequently, observations at 3.5mm are the longest wavelengths with active VLBI for which the intrinsic structure is not sub-dominant to scattering (VLBI observations of Sgr A* at wavelengths between 3.5mm and 7mm are very difficult because of atmospheric oxygen absorption).

As discussed by Psaltis et al. [2018] and J18, the $\lambda^2$ and Gaussian scattering behavior of Sgr A* are universally expected if 1) the intrinsic source size $\theta_{\text{src}}$ is subdominant to the scatter broadening angle $\theta_{\text{scatt}}$, and 2) the diffractive scale of the scattering $r_{\text{diff}} \sim \lambda / \theta_{\text{scatt}}$ is smaller than the dissipation scale of turbulence in the scattering material. Thus, even though the angular broadening size and shape are measured very precisely for Sgr A* at centimeter wavelengths, the constraints on the overall scattering properties are quite weak. The expected dissipation scale in the ISM is $10^2$ – $10^3$ km [e.g., Spangler & Gwinn, 1990], so the expected transition to non-$\lambda^2$ and non-Gaussian scattering (i.e., when the dissipation scale is comparable to the diffractive scale) for Sgr A* occurs at wavelengths of a few millimeters. Consequently, the scattering properties of Sgr A* measured at centimeter wavelengths cannot be confidently extrapolated to millimeter wavelengths. The uncertainties can be parameterized using physical models for the scattering material, which typically invoke an anisotropic power-law for the power spectrum of phase fluctuations, with the power-law extending between a maximum scale (the outer scale $r_{\text{out}}$) and a minimum scale (the inner scale $r_{\text{in}}$). In such a generalization, the scattering properties depend on a spectral index $\alpha$, and on the inner scale of the turbulence, $r_{\text{in}}$. In this paper, we use the scattering model presented in Psaltis et al. [2018] with parameters for Sgr A* determined by J18.

The discovery by Gwinn et al. [2014] of scattering-induced substructure in images of Sgr A* at 1.3cm gives an additional constraint on the scattering properties of Sgr A*. This substructure is caused by modes in the scattering material on scales comparable to the image extent (much larger than $r_{\text{diff}}$), so scattering models with identical scatter-broadening may still exhibit strong differences in their scattering substructure. The substructure manifests in the visibility domain as "refractive noise", which is an additive complex noise component with broad correlation structure across baselines and time [Johnson & Narayan, 2016]. Using observations of Sgr A* from 1.3 mm to 30 cm, J18 have shown that the combined image broadening and substructure strongly constrains the power spectrum of density fluctuations. However, a degeneracy between $\alpha$ and $r_{\text{in}}$ persists, and extrapolating the strength of refractive effects to millimeter wavelengths is still quite uncertain.

Two scattering models effectively bracket the range of possibilities for Sgr A*. One model (hereafter J18) has a power-law spectral index $\alpha = 1.38$ (near the expected value for 3D Kolmogorov turbulence, $\alpha = 5/3$) and $r_{\text{in}} = 800$ km (near the expected ion gyroradius in the ionized ISM). The second is motivated by Goldreich & Sridhar [2006, hereafter GS06], who proposed that the scattering of Sgr A* could be caused by thin current sheets in the ISM; it has $\alpha = 0$ and $r_{\text{in}} \sim 2 \times 10^6$ km. The inner scale in this latter model is several orders of magnitude larger
than originally proposed by GS06, but this larger value is required to produce the refractive noise observed at 1.3 and 3.5 cm. Both the J18 and GS06 models are consistent with all existing measurements of the angular broadening of Sgr A* and with the refractive noise at centimeter wavelengths, but the GS06 model would produce more refractive noise than the J18 model on long baselines at 3.5 mm, with even more pronounced enhancement for EHT observations [by roughly an order of magnitude; see Zhu et al., 2019]. While long-baseline measurements at 3.5 mm can discriminate between these possibilities, observations to-date have been inadequate for an unambiguous detection of refractive substructure at this wavelength [O16; B16; B19]. New observations with ALMA joining 3.5 mm VLBI, with unprecedented resolution and sensitivity, give the opportunity for long-baseline detections of refractive noise at millimeter wavelengths that can enable discrimination between the two scattering models.

2.3 Observations and data reduction

Observations of Sgr A* (αJ2000 = 17h45m40s.0361, δJ2000 = −29°00′28″.168) were made with the GMVA, composed of the eight Very Long Baseline Array (VLBA) antennas equipped with 86 GHz receivers, the Green Bank Telescope (GB), the Yebes 40-m telescope (YS), the IRAM 30-m telescope (PV), the Effelsberg 100-m telescope (EB), and the ALMA phased array (AA) consisting of 37 phased antennas. The observations were conducted on 3 April 2017 as part of the first offered VLBI session with ALMA (project code MB007). We recorded a total bandwidth of 256 MHz per polarization divided in 4 intermediate frequencies (IFs) of 116 channels each. The 12 h track (4 h with the European sub-array and 8 h with ALMA) included three calibrator sources: 1749+096, NRAO 530, and J1924−2914. The total integration time on Sgr A* with ALMA was 5.76 h.

The data were processed with the VLBI correlator at the Max Planck Institute for Radio Astronomy using DiFX [Deller et al., 2011]. After correlation, reduction was carried out using the Haystack Observatory Postprocessing System1 (HOPS) supported by a suite of auxiliary calibration scripts presented in Blackburn et al. [2019], with additional validation and cross-checks from the NRAO Astronomical Image Processing System [AIPS; Greisen, 2003]. The HOPS software package in its current form arose out of the development of the Mark IV VLBI Correlator, see Whitney et al. [2004]. During the HOPS reduction, ALMA baselines were used to estimate stable instrumental phase bandpass and delay between right and left circular polarization relative to the other stations. ALMA or GBT baselines (depending on signal-to-noise) were used to remove stochastic differential atmospheric phase within a scan. Because atmospheric phase corrections are required on short (~second) timescales, leading to a large number of free parameters to fit, a round-robin calibration was used to avoid self-tuning: baseline visibility phases on each 58 MHz IF were estimated using only the remaining 3 IFs, which have independent thermal noise. The integration time for rapid phase corrections was automatically chosen by balancing errors from random thermal variation to those due to atmospheric phase drift, and thus varied with the available signal-to-noise. The median effective integration time was 4.5 seconds. During a final stage of reduction with the HOPS fringe fitter fourfit, fringe solutions for each scan were fixed to

1https://www.haystack.mit.edu/tech/vlbi/hops.html
2.3 Observations and data reduction

Figure 2.1: Top: The \((u,v)\)-coverage of Sgr A*\(^*\). Each symbol denotes a scan-averaged measurement: filled colored circles are strong detections; hollow colored circles are weak detections (constrained fringe delay and rate but signal-to-noise ratio (SNR) < 6); and hollow gray circles are non-detections (unconstrained fringe delay and rate) after processing through HOPS. Bottom: The SNR for scan-averaged visibilities on Sgr A*\(^*\) as a function of projected baseline length, showing only detections. All detections beyond \(\sim 1\) GA are on baselines to ALMA.
Chapter 2: The Size, Shape, and Scattering of Sgr A*  

Figure 2.2: Left: the \((u,v)\)-coverage of NRAO 530 (symbols are as defined in Figure 2.1). Right: closure-only image of NRAO 530 using the eht-imaging library [Chael et al., 2018b], the contour levels start from 1.2% of the peak and increase in factors of two. The observations have a uniform-weighted beam = \((111 \times 83) \ \mu\text{as}, \ PA = 32^\circ\).

We performed a-priori amplitude calibration using provided telescope gain information and measured system temperatures during the observations. The heterogeneity of the stations in the GMVA required us to adopt a careful approach to the amplitude calibration. The calibration for ALMA was fully provided by the ALMA quality assurance (QA2) team [Goddi et al., 2019], and system equivalent flux densities (SEFDs) were generated with a high time cadence by PolConvert [Martí-Vidal et al., 2016b]. Both YS and PV measure effective system temperatures via the chopper wheel method, and thus do not require an additional opacity correction to their SEFDs. However, the rest of the array (VLBA, GB, EB) measures system temperatures via the noise diode method, requiring an additional opacity correction to account for atmospheric attenuation of the visibility amplitudes. Unfortunately, several VLBA stations observed in difficult weather conditions (ice, wind, rain), leading to limited detections on baselines to Owens Valley (OV), North Liberty (NL) and Pie Town (PT) stations. Additionally, observations at PV suffered from phase coherence losses in the signal chain during the observations, leading to poor quality data and lower visibility amplitudes on those baselines, which cannot be rescaled with a-priori calibration information. Figure 2.1 shows the detections and non-detections for Sgr A* (top panel) and corresponding signal-to-noise ratio of scan-averaged visibilities for Sgr A* detections. All detections beyond ~1\(\Gamma\lambda\) are on baselines to ALMA. After a-priori calibration, we can proceed with imaging routines to determine the morphology of the calibrators and the target source.
2.4 Imaging

We employ the eht-imaging library\(^2\), a regularized maximum likelihood imaging software package, to image our sources [Chael et al., 2016, 2018b]. Due to the elevated noise level for the VLBA in our observations and the scattering properties of Sagittarius A*, standard imaging software packages like AIPS [Greisen, 2003] or Difmap [Shepherd et al., 1995] do not offer the flexibility and necessary tools to obtain an unscattered image of the source. The eht-imaging library is a Python-based software package that is easily scriptable, flexible and modular. It is able to make images with various data products (closure phase and amplitude, bispectra, visibilities), and it contains a suite of image “regularizers” such as maximum entropy and sparsity regularization. The library also possesses a routine for “stochastic optics”, a regularized implementation of scattering mitigation presented in Johnson [2016], making it a natural choice for our analysis. In this section we present our imaging methods for both calibrators (Section 2.4.1) and for Sgr A* (Section 2.4.2).

2.4.1 Calibrators NRAO 530 and J1924−2914

Both NRAO 530 and J1924−2914 appear point-like to ALMA when acting as a connected-element interferometer (∼70 kλ, ∼3 arcseconds resolution), with NRAO 530 having a flux density of 2.8 ± 0.3 Jy and J1924−2914 having a flux density of 5.0 ± 0.5 Jy (as measured by interferometric-ALMA). Even on the angular scales probed by VLBI, both sources are very compact and stable, making them ideal for imaging. The operational difficulties and poor weather conditions at the VLBA were largely offset by the high sensitivity of ALMA. The extent of all detections is shown in the left panel of Figure 2.2 for NRAO 530 and Figure 2.3 for J1924−2914. A third calibrator

\(^2\)https://github.com/achael/eht-imaging
Table 2.1: Station median multiplicative gains to the visibility amplitudes.

<table>
<thead>
<tr>
<th>Station</th>
<th>Sgr A*</th>
<th>NRAO 530</th>
<th>J1924–2914</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR</td>
<td>$2.2^{+1.5}_{-0.8}$</td>
<td>$1.7^{+0.5}_{-0.5}$</td>
<td>$2.0^{+1.4}_{-1.8}$</td>
</tr>
<tr>
<td>FD</td>
<td>$2.2^{+1.2}_{-0.6}$</td>
<td>$1.9^{+1.2}_{-0.5}$</td>
<td>$1.8^{+0.7}_{-0.4}$</td>
</tr>
<tr>
<td>GB</td>
<td>$1.2^{+1.7}_{-0.4}$</td>
<td>$1.1^{+0.5}_{-0.1}$</td>
<td>$1.2^{+0.7}_{-0.4}$</td>
</tr>
<tr>
<td>KP</td>
<td>$2.4^{+2.2}_{-0.6}$</td>
<td>$2.2^{+1.2}_{-0.4}$</td>
<td>$2.1^{+0.7}_{-0.4}$</td>
</tr>
<tr>
<td>LA</td>
<td>$2.2^{+2.8}_{-1.0}$</td>
<td>$1.9^{+1.7}_{-0.7}$</td>
<td>$2.9^{+2.0}_{-2.1}$</td>
</tr>
<tr>
<td>NL</td>
<td>$4.6^{+13.3}_{-2.1}$</td>
<td>$4.7^{+9.7}_{-1.5}$</td>
<td>$5.0^{+22.6}_{-2.3}$</td>
</tr>
<tr>
<td>OV</td>
<td>$1.9^{+3.1}_{-1.0}$</td>
<td>$1.9^{+0.9}_{-0.6}$</td>
<td>$1.7^{+1.6}_{-0.3}$</td>
</tr>
<tr>
<td>PT</td>
<td>$11.4^{+2.2}_{-5.3}$</td>
<td>$19.3^{+17.8}_{-13.3}$</td>
<td>$12.9^{+36.7}_{-8.4}$</td>
</tr>
</tbody>
</table>

NOTE—Median (and 95th percentile) multiplicative gains to the visibility amplitudes for common stations from the two calibration methods: 1) self-calibration of Sgr A* amplitudes below 0.75Gλ to the Gaussian source estimated from O16; B19, and 2) self-calibration of NRAO 530 and J1924–2914 observations to the images produced with closure phases and closure amplitudes. The European stations and ALMA are not shown as they are not self-calibrated for all three sources. We flagged NL and PT due to their high median gain and erratic gain solutions.

was also observed, 1749+096, but only for a few minutes with the full array, and is thus omitted from further analysis.

The large number of detections on both NRAO 530 and J1924–2914 led to a correspondingly large number of closure phases and closure amplitudes. We thus imaged both sources using only closure quantities, following the method from Chael et al. [2018b], constraining the total flux of the image to match measurements from interferometric-ALMA. We present images of the two calibrators in Figure 2.2 and Figure 2.3 (right panels). The morphology of NRAO 530 is consistent with previous observations of the source [Bower et al., 1997; Bower & Backer, 1998; Feng et al., 2006; Chen et al., 2010; Lu et al., 2011c]. The elongation of the J1924–2914 jet in the north-west direction at 86 GHz is consistent with mm-jet studies from previous observations at 43 GHz by Shen et al. [2002] and 230 GHz by Lu et al. [2012]. These two sources are common calibrators for Sgr A*. They are therefore particularly useful to study at multiple frequencies to adequately calibrate observations at 1.3 mm from the EHT.

2.4.2 Sagittarius A*

Self-calibration

We obtained far fewer detections on Sgr A* than on the calibrators, and our detections also had lower signal-to-noise ratio (SNR). Consequently, we did not have enough information to synthesize images of Sgr A* using only closure quantities. Moreover, due to the suboptimal performance of the VLBA (bad weather, signal loss likely from pointing issues), additional amplitude calibration was necessary to mitigate severe signal losses at various stations.

We utilized two methods for amplitude calibration:
Figure 2.4: Noise-debiased correlated flux density of Sgr A* as a function of projected baseline length for data after self-calibrating to the Gaussian source from O16 and B19 using only baselines shorter than 0.75 Gλ. Because the a-priori calibration for the GBT was excellent (see Table 2.1), we did not apply the derived GBT gains. Dashed dark blue curves show expected visibilities along the major and minor axes for an anisotropic Gaussian source with FWHM of 215 μas by 140 μas (the source size from O16 and B19). All detections beyond ∼1Gλ are baselines to ALMA, and all show marked deviations from the Gaussian curves.
1. We self-calibrated to closure-only images of NRAO 530 and J1924–2914 to obtain smoothed station gain trends.

2. We self-calibrated all Sgr A* visibility amplitudes within 0.75Gλ (predominantly intra-VLBA measurements) using an anisotropic Gaussian visibility function determined by previous 3.5 mm experiments [O16; B19], with the total flux set by the interferometric-ALMA measurement.

For the second method, we used a visibility function corresponding to a Gaussian source size of 215 by 140 µas with a position angle of 80° (east of north) and a total flux density of 2.0±0.2 Jy. The choice of the Gaussian size is motivated by similar results obtained for O16 and B19 taken one month apart, showing stable source dimensions. Both these experiments had the high sensitivity of the LMT, adding north-south coverage to recover the minor axis size with greater accuracy than older experiments. In our interferometric-ALMA measurements, Sgr A* has flux density variations at the 10% level on a timescale of about 4 hours, not significantly affecting our static imaging. Note that gains were derived by self-calibration using only short baselines, but because they are station-based, they were then applied to correct visibility amplitudes on longer baselines as well.

The two methods gave comparable gain solutions, hence validating the Gaussian assumption for short-baseline measurements (Table 2.1 shows median multiplicative station gains to the visibility amplitudes). We flagged the VLBA stations NL and PT, which showed extreme signal loss in both methods. The GBT performed well for all three sources, so we chose to keep the original a-priori calibration. Because GBT is only linked to NL in the inner 0.75 Gλ baseline cut for Sgr A*, the derived gains for GBT introduce large variations to the ALMA-GBT amplitudes that come from difficulty locking NL gains due to its bad weather. Ignoring the self-calibration solutions gave more stable amplitudes on the ALMA-GBT baseline.

Figure 2.4 shows the scan-averaged visibilities for Sgr A* after self-calibration of the inner 0.75 Gλ baselines to the Gaussian source size (method 2). All detections above 1 Gλ are new measurements to ALMA. The ALMA-GBT baseline has significantly higher flux density than expected from the minor axis of the previously fitted Gaussian source size from O16 and B19. VLBA detections to ALMA show clear deviations from Gaussian behavior.

**Imaging with regularized maximum likelihood**

The performance of the VLBA impaired our ability to model-fit to the dataset and obtain an accurate source size estimate using only short baselines (i.e., baselines that do not heavily resolve the source). In addition, large measurement uncertainties for the visibility amplitudes on intra-VLBA baselines made image convergence difficult and unstable. We thus implemented a new imaging regularization: we constrained the second central moment of the image to match more robust measurements of the scattered source size from Gaussian model fitting to previous observations [O16; B19]. If we think of the centroid (first moment) of the image as the mean position of the emission, its variance (or second moment) is the spread of emission from the mean, equivalent to the extent of the source along its principal axes [Hu, 1962]. The regularization is equivalent to constraining the curvature of the visibility function at zero baseline. This method helps to calibrate short-baseline visibilities during the imaging process, while allowing long-baseline de-
2.4 Imaging

Figure 2.5: Left: the scattered image of Sgr A*, reconstructed with the second moment regularizer and stochastic optics ($\theta_{\text{maj}} = 228 \pm 46 \ \mu\text{as}, \ \theta_{\text{min}} = 143 \pm 20 \ \mu\text{as}$ from LSQ). Right: the reconstructed image from stochastic optics [Johnson, 2016] of the intrinsic source ($\theta_{\text{maj}} = 120 \pm 34 \ \mu\text{as}, \ \theta_{\text{min}} = 100 \pm 18 \ \mu\text{as}$ from LSQ). In each panel, the ellipses at the bottom indicate half the size of the scatter-broadening kernel ($\theta_{\text{maj}} = 159.9 \ \mu\text{as}, \ \theta_{\text{min}} = 79.5 \ \mu\text{as}, \ \text{PA} = 81.9^\circ$) and of the observing beam.

Sections to ALMA to still recover smaller scale structure in our images. This method is now included and implemented in the eht-imaging library via gradient descent minimization (the effects and fidelity of the regularizer will be presented in Issaoun et al. [2019a]). We also made use of the “stochastic optics” scattering mitigation code from Johnson [2016] to disentangle the effects of scattering and produce the intrinsic image of Sgr A*.

To reach our final result, we first imaged the scattered source using closure quantities and visibility amplitudes (with equal weights). The regularizers used in the scattered image, with a weighting of 10% of the data weights, were: Gull-Skilling maximum entropy; total squared variation; and second moment regularization with the second moment matrix given by that of the Gaussian used for self-calibration. Each of these regularizers favors particular image features, while enforcing image positivity and a total flux density constraint. Gull-Skilling entropy favors pixel-by-pixel similarity to the prior image (we used the previously fitted Gaussian source as the prior). Total-squared variation regularization favors small image gradients, producing smooth edges (see Chael et al. [2018b] for a detailed discussion of these regularizers). Second moment regularization constrains the second derivative of the visibility function at the zero baseline (which is proportional to the second central image moment) to match a specified value; we thereby constrained our short baselines to match those of the Gaussian source measured in previous experiments [O16; B19] without imposing assumptions on the visibilities measured by longer baselines, which reflect image substructure. In the scattering mitigation code, the second moment regularization is only applied to the observed image, such that the intrinsic image derived by the scattering deconvolution is not directly constrained by the regularizer but still remains within physical size ranges. After imaging with closure quantities and corrected visibility amplitudes, we then self-calibrated the visibility phases and amplitudes to the obtained scattered image before imaging with stochastic optics (using the same regularization parameters).
Figure 2.6: Model and reconstructed images from four example 3D GRMHD models, plotted here in linear scale. The contour levels represent 25, 50 and 75% of the peak flux. The first column shows the original model images as given from simulations: “th+κ disk” is a thermal disk model with 1% accelerated particles in a power-law (κ) distribution; “th jet” is a thermal jet model; “th+κ jet” is a thermal jet model with 10% accelerated particles in a κ distribution [Mościbrodzka et al., 2009, 2014, 2016; Davelaar et al., 2018]. The inclinations of the models are given in the parentheses. The second column shows the model images scattered with the J18 scattering model: these are the images sampled to make the simulated observations. The third column shows the observed (scattered) image reconstructed with the second moment regularizer and stochastic optics, and the fourth column shows the reconstructed image from stochastic optics of the corresponding intrinsic source. In the third and fourth columns, the ellipses at the bottom indicate half the size of the scatter-broadening kernel and of the observing beam.
The stochastic optics framework is implemented within the `eht-imaging` library via regularized maximum likelihood. The code solves for the unscattered image by identifying, separating and mitigating the two main components of the scattering screen, introduced in Section 2.2: small-scale diffractive modes that blur the image, causing the ensemble-average scattered image to be a convolution of the true image and the scattering kernel (predominantly east-west scatter-broadening); and large-scale refractive modes that introduce stochastic image substructure (ripples distorting the image). The code simultaneously solves for the unscattered image and the large-scale phase screen causing refractive scattering, while assuming a given model for the diffractive blurring kernel and the refractive power spectrum $Q(q)$ (governing the time-averaged scattering properties). In our case, we used the scattering kernel (with a size of $(159.9 \times 79.5) \mu$as, PA of $81.9^\circ$) and power spectrum (with $\alpha = 1.38$ and $r_{in} = 800$ km) from the J18 scattering model. See Johnson [2016] for a more detailed description of the method. Two iterations of stochastic imaging and self-calibration are done for convergence. We present in Figure 2.5 our resulting intrinsic and scattered images of Sgr A*. 
### Table 2.2: Comparison of the true size and the derived size from imaging from synthetic datasets for four simulated images.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>$\theta_{\text{maj}}$ (µas)</th>
<th>$\theta_{\text{min}}$ (µas)</th>
<th>Axial ratio</th>
<th>PA (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>121.0</td>
<td>97.3</td>
<td>1.24</td>
<td>105.4</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>184.0</td>
<td>131.8</td>
<td>1.4</td>
<td>87.6</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>63.0 (0.4θ_{\text{beam}})</td>
<td>34.5 (0.1θ_{\text{beam}})</td>
<td>0.16 (0.1θ_{\text{beam}})</td>
<td>17.8 (0.6θ_{\text{beam}})</td>
</tr>
<tr>
<td>Th+κ disk (60°)</td>
<td>True</td>
<td>79.7</td>
<td>77.0</td>
<td>1.04</td>
<td>109.9</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>101.9</td>
<td>59.6</td>
<td>1.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>22.2 (0.1θ_{\text{beam}})</td>
<td>17.4 (0.1θ_{\text{beam}})</td>
<td>0.66 (0.1θ_{\text{beam}})</td>
<td>69.3 (≥ 0.4θ_{\text{beam}})</td>
</tr>
<tr>
<td>LSQ</td>
<td>True</td>
<td>112.5</td>
<td>99.0</td>
<td>1.14</td>
<td>13.8</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>148.7</td>
<td>124.8</td>
<td>1.19</td>
<td>74.2</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>36.2 (0.3θ_{\text{beam}})</td>
<td>25.8 (0.1θ_{\text{beam}})</td>
<td>0.05 (0.02θ_{\text{beam}})</td>
<td>60.4 (≥ θ_{\text{beam}})</td>
</tr>
<tr>
<td>Th jet (5°)</td>
<td>True</td>
<td>88.0</td>
<td>81.2</td>
<td>1.08</td>
<td>179.7</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>65.5</td>
<td>51.9</td>
<td>1.26</td>
<td>158.4</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>22.5 (0.2θ_{\text{beam}})</td>
<td>29.3 (0.1θ_{\text{beam}})</td>
<td>0.18 (0.03θ_{\text{beam}})</td>
<td>21.3 (0.2θ_{\text{beam}})</td>
</tr>
<tr>
<td>LSQ</td>
<td>True</td>
<td>174.0</td>
<td>65.8</td>
<td>2.64</td>
<td>179.8</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>178.1</td>
<td>135.3</td>
<td>1.32</td>
<td>176.4</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>4.1 (0.02θ_{\text{beam}})</td>
<td>69.5 (0.5θ_{\text{beam}})</td>
<td>1.32 (0.06θ_{\text{beam}})</td>
<td>3.4 (0.3θ_{\text{beam}})</td>
</tr>
<tr>
<td>Th jet (90°)</td>
<td>True</td>
<td>160.8</td>
<td>63.2</td>
<td>2.54</td>
<td>178.5</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>130.3</td>
<td>42.4</td>
<td>3.07</td>
<td>177.1</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>30.5 (0.2θ_{\text{beam}})</td>
<td>20.8 (0.1θ_{\text{beam}})</td>
<td>0.53 (0.04θ_{\text{beam}})</td>
<td>1.7 (0.2θ_{\text{beam}})</td>
</tr>
<tr>
<td>LSQ</td>
<td>True</td>
<td>182.4</td>
<td>65.7</td>
<td>2.78</td>
<td>179.7</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>177.5</td>
<td>127.6</td>
<td>1.4</td>
<td>177.6</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>4.9 (0.02θ_{\text{beam}})</td>
<td>61.9 (0.4θ_{\text{beam}})</td>
<td>1.38 (0.06θ_{\text{beam}})</td>
<td>2.1 (0.2θ_{\text{beam}})</td>
</tr>
<tr>
<td>Th+κ jet (90°)</td>
<td>True</td>
<td>166.6</td>
<td>62.9</td>
<td>2.65</td>
<td>178.7</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>141.5</td>
<td>49.9</td>
<td>2.83</td>
<td>179.2</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>25.1 (0.1θ_{\text{beam}})</td>
<td>13.0 (± 0.1θ_{\text{beam}})</td>
<td>0.18 (0.02θ_{\text{beam}})</td>
<td>0.5 (0.1θ_{\text{beam}})</td>
</tr>
</tbody>
</table>

Note. — In each case, we compute the sizes using two methods: directly from the image second central moment ("2nd mom."), and from a 2D Gaussian fit to the image with least-squares minimization ("LSQ"). We give the absolute difference between the true and estimated values and also express the difference as a fraction of the projected beam FWHM θ_{beam} along the measured axis, or as the fraction of the propagated error from the beam-widths on both axes for the axial ratio. The uncertainty on the position angle (PA) is expressed as the fraction of one-dimensional beam blurring of the image for which the standard deviation in PA with blurring along different directions matches the difference between the true and measured PA (see text for additional details).
Uncertainties of image-derived parameters

To determine the uncertainties in the imaging method and size measurements for Sgr A*, we performed imaging tests on simulated observations where the intrinsic model image was known. We tested our imaging method on four snapshots from 3D GRMHD simulations of Sgr A* at 86 GHz [Mościbrodzka et al., 2009, 2014, 2016; Davelaar et al., 2018], using the same sampling, coverage and noise as our observations. The model images were scattered with the J18 scattering model and sampled with our GMVA+ALMA coverage, before being imaged via the same imaging routine applied to the Sgr A* data described above.

While the imaging procedure is identical, these reconstructions do have some advantages relative to our reconstruction of the actual observations. For example, we used the ensemble-average properties of the J18 scattering model as inputs to the scattering mitigation: i.e., we assume perfect knowledge of the diffractive scattering kernel and the time-averaged power spectrum. We also measure the second moment of the scattered simulated images and use it as an exact input to the second moment regularization. Because the scattering is subdominant to intrinsic structure and because the second moment is estimated to excellent accuracy in previous experiments, we do not expect either of these effects to significantly advantage the reconstructions of simulated data.

In Figure 2.6 we present the original 3D GRMHD model images, the model images scattered with the J18 scattering model (as observed in the simulated observations), and the reconstructed observed (scattered) and intrinsic images from the imaging method. In Table 2.2 we compare the true intrinsic source sizes from the models to the intrinsic source sizes derived from the imaging routine. We determined the source size parameters using two methods: first by measuring the second central moment of the image (2nd mom.) and deriving Gaussian parameters; and second by doing a 2D Gaussian fit with a least-squares minimization (LSQ) onto the image.
### Table 2.3: Observed and intrinsic sizes for Sgr A* at 86 GHz.

<table>
<thead>
<tr>
<th>Reference</th>
<th>$\theta_{\text{maj,obs}}$ (µas)</th>
<th>$\theta_{\text{min,obs}}$ (µas)</th>
<th>PA$_{\text{obs}}$ (deg)</th>
<th>Axial ratio</th>
<th>$\theta_{\text{maj,int}}$ (µas)</th>
<th>$\theta_{\text{min,int}}$ (µas)</th>
<th>PA$_{\text{int}}$ (deg)</th>
<th>Axial ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rogers et al. [1994]</td>
<td>150 ± 50</td>
<td>-</td>
<td>-</td>
<td>&lt; 130</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Krichbaum et al. [1998]</td>
<td>190 ± 30</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Doeleman et al. [2001]</td>
<td>180 ± 20</td>
<td>-</td>
<td>-</td>
<td>&lt; 130</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shen et al. [2005]</td>
<td>$210^{+20}_{-10}$</td>
<td>$130^{+50}_{-13}$</td>
<td>$79^{+12}_{-33}$</td>
<td>1.6$^{+0.4}_{-0.5}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lu et al. [2011b]</td>
<td>210 ± 10</td>
<td>130 ± 10</td>
<td>83 ± 2</td>
<td>1.6 ± 0.1</td>
<td>139 ± 17</td>
<td>102 ± 21</td>
<td>-</td>
<td>1.4 ± 0.3</td>
</tr>
<tr>
<td>O16 BD183C</td>
<td>213 ± 2</td>
<td>138 ± 4</td>
<td>81 ± 2</td>
<td>1.54 ± 0.04</td>
<td>142 ± 9</td>
<td>114 ± 15</td>
<td>-</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>O16 BD183D</td>
<td>222 ± 4</td>
<td>146 ± 4</td>
<td>75 ± 3</td>
<td>1.52 ± 0.05</td>
<td>155 ± 9</td>
<td>122 ± 14</td>
<td>-</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>B19 (clos.amp.)</td>
<td>215.1 ± 0.4</td>
<td>145 ± 2</td>
<td>77.9 ± 0.4</td>
<td>1.48 ± 0.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B19 (selfcal)</td>
<td>217 ± 22</td>
<td>165 ± 17</td>
<td>77 ± 15</td>
<td>1.3 ± 0.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>J18 BD183C</td>
<td>215 ± 4</td>
<td>139 ± 4</td>
<td>81 ± 3</td>
<td>1.55 ± 0.05</td>
<td>143$^{+11}_{-12}$</td>
<td>114$^{+7}_{-8}$</td>
<td>-</td>
<td>1.25$^{+0.20}_{-0.16}$</td>
</tr>
<tr>
<td>This work ($2^{nd}$ mom.)$^a$</td>
<td>239 ± 57</td>
<td>172 ± 103</td>
<td>84 ± 2</td>
<td>1.4$^{+1.1}_{-0.4}$</td>
<td>176 ± 57</td>
<td>152 ± 103</td>
<td>85.2 ± 44$^b$</td>
<td>1.2$^{+1.1}_{-0.2}$</td>
</tr>
<tr>
<td>This work (LSQ)$^a$</td>
<td>228 ± 46</td>
<td>143 ± 20</td>
<td>86 ± 2</td>
<td>1.6 ± 0.3</td>
<td>120 ± 34</td>
<td>100 ± 18</td>
<td>96.0 ± 32$^b$</td>
<td>1.2$^{+0.3}_{-0.2}$</td>
</tr>
</tbody>
</table>

Notes. —

$^a$Image-domain size estimates. The stated uncertainties are derived using the largest parameter errors for reconstructions of simulated images.

$^b$Position angle estimates are not meaningfully constrained because of the near symmetry of the major/minor axes.
Next, we evaluate the difference between true and reconstructed image parameters. We sought to define an approach that quantifies these differences in a way that is related to the reconstructed image properties and the observing beam. When expressed in this way, we can use parameter errors on these reconstructed simulated images to predict uncertainties on parameters derived from our reconstructed image with data.

To this end, Table 2.2 expresses the difference between the true and measured source major and minor axes as a fraction of the projected beam FWHM $\theta_{\text{beam}}$ along the corresponding axis. For the axial ratio, we express the difference between the true and measured ratios as a fraction of the cumulative error from both axes (the projected beam-widths along the measured major and minor axes added quadratically).

However, while it is straightforward and well-motivated to express uncertainties on axis lengths and their ratio in terms of the observing beam, uncertainty on the position angle (PA) is more subtle. We opted to create an ensemble of beam-convolved reconstructed images and to use the scatter in the PA of the ensemble as an estimate of the PA uncertainty. The ensemble of images is constructed by convolving the single reconstructed image with an ensemble of narrow beams, sampling all position angles. Each of these beams has a major axis size given by the projected observing beam size along the same position angle and a minor axis size of zero. We thereby stretch the image along each direction, up to the extent of the observing beam, and examine the overall dependence of the reconstructed image on this stretching. With this approach, images that are nearly isotropic will have large PA uncertainty, while highly elongated images (relative to the beam size) will have small PA uncertainty.

In general, we find that the LSQ method fares better than 2nd moment for determining the source parameters, likely due to weak extended flux in the images skewing the second moment parameters to larger values. As expected, both methods perform poorly when determining the position angle of a fairly symmetrical source, for which it remains largely unconstrained. However, for more elongated source geometry, both methods are able to accurately recover the intrinsic position angle. We adopt the LSQ method to quantify the size of Sgr A* via image-domain fitting. Although the Gaussian approximation does not describe fully our source morphology, it is suitable for comparisons to visibility-domain model fits from previous observations of Sgr A* presented in Section 2.5.

2.5 Results

2.5.1 Intrinsic source constraints from imaging

Figure 2.5 shows the unscattered and scattered images of Sgr A*, as imaged following the method described in Section 2.4. The (uniform-weighted) beam size of the Sgr A* observations is $(235 \times 87) \mu\text{as}$, with a position angle (east of north) of $53.6^\circ$. While the shorter baselines of the array (intra-VLBA, VLBA-GBT, intra-European) see primarily a Gaussian source elongated in the east-west direction, longer baselines are expected to pick up on non-Gaussian source structure or refractive noise from interstellar scattering. In this particular observation, our longest baselines are mainly north-south to ALMA (see Figure 2.1), where scattering has less of an effect on the source. As seen in Figure 2.5, left panel, the reconstructed scattered image looks very smooth and Gaussian-like,
showing no obvious refractive noise in the image. We also see a similar outcome in our imaging tests, presented in Section 2.4.2. Although the scattered images (second column in Figure 2.6) have visible ripples of scattering substructure, the reconstructed scattered images (third column) appear very smooth. This is likely because our GMVA+ALMA observations sample low levels of refractive noise mainly along the north-south direction, whereas our east-west sensitivity and resolution do not provide adequate detections of scattering substructure to be able to reconstruct the fine structure in the scattered images. Thus the low level of refractive noise detected on our ALMA baselines does not produce visible distortions in the reconstructed scattered image.

We present the measured source sizes using our two methods (2nd mom. and LSQ) in Table 2.3, along with historical measurements and estimates. The uncertainties are conservative estimates taken from the largest relative uncertainties on the parameters of simulated images for each method (see Table 2.2). We assume a Gaussian source geometry for size estimates and comparisons, but this may not be the correct source model. As seen in the example images (Figure 2.6), the true and reconstructed intrinsic images are not Gaussian, therefore this choice of parametrization is only to simplify comparisons with previous measurements and simulations. We find that our source size measurements are consistent with previous observations and indicate the source dimensions and small asymmetry are persistent across multiple years.

Lastly, we note that uncertainties in the intrinsic size caused by remaining uncertainties in the scattering kernel are quite small ($\lesssim 10$ μas), even allowing for the full range of uncertainty on $\alpha$ and $r_{\text{in}}$ [J18]. The reason they are small is because the scattering parameters for angular broadening are estimated to an accuracy of a few percent, and because the intrinsic structure is not subdominant to scatter broadening.

### 2.5.2 Intrinsic source constraints from closure phases

Closure phases provide an alternative and complementary assessment of source asymmetry directly from observations. They are weakly affected by refractive scattering and are unaffected by station-based calibration issues. Thus, they offer robust information on the intrinsic properties of Sgr A*.

We computed closure phases for all sources from scan-averaged visibilities. The GMVA+ALMA array contains 13 stations, yielding many triangles with a wide range of sizes. As seen in Figures 2.2 and 2.3, there are multiple long-baseline detections to ALMA on calibrators that do not appear for Sgr A* (Figure 2.1). We thus selected three example triangles of different sizes and orientations that are present for the two main calibrator sources (NRAO 530 and J1924−2914) and with multiple detections for Sgr A*.

We present in Figure 2.7 the closure phases on three representative triangles: a small intra-VLBA (LA-KP-FD) triangle; an east-west medium-sized triangle to GBT (GB-KP-FD); and a long north-south triangle to ALMA (AA-GB-FD). Although all three triangles provide robust detections for all three sources, with non-zero closure phases for the calibrators, Sgr A* closure phases remain very close to zero: the weighted mean closure phase on AA-GB-FD is $-1.1 \pm 2.4^\circ$; the weighted mean closure phase on GB-KP-FD is $-1.7 \pm 1.1^\circ$; and the weighted mean closure phase on LA-KP-FD is $-1.8 \pm 1.1^\circ$. The largest closure phases on all three triangles deviate from zero by less than $3\sigma$.

O16 and B16 detected small non-zero closure phases ($\lesssim 10^\circ$) on triangles including the highly...
Figure 2.7: Scan-averaged closure phases for Sgr A*, NRAO 530 and J1924–2914 on three triangles (LA-KP-FD, GBT-KP-FD, ALMA-GBT-FD) formed after processing through HOPS. The larger uncertainties on the ALMA-GBT-FD triangle are primarily because of low correlated flux density on the ALMA-FD baseline (see Figure 2.4). Non-zero closure phase indicates source asymmetry. Although NRAO 530 and J1924–2914 show significant deviations from zero, all Sgr A* closure phases are consistent with zero within 3σ.
sensitive LMT and/or GBT. These non-zero closure phases were observed on triangles not present in our GMVA+ALMA observations, and they probed different scales and directions from our new predominantly north-south triangles with ALMA. Deviations of a few degrees, as observed by O16 and B16, fall within our confidence bounds due to low signal-to-noise on VLBA baselines, and thus would not be detectable with our current observations. Moreover, the geometrical models to describe the asymmetry in B16 produce closure phases on our triangles that would be indistinguishable from zero with our current measurements. Thus, our results are consistent with previous observations of Sgr A*.

2.6 Discussion

2.6.1 Constraints on the Refractive Scattering of Sgr A*

Our longest baselines heavily resolve the scattered image of Sgr A* while also providing exceptional sensitivity (especially baselines to ALMA). Therefore they are sensitive to a non-Gaussian scattering kernel (from a finite inner scale) and to “refractive noise,” which corresponds to image substructure introduced by interstellar scattering. In this section, we use our long-baseline measurements to constrain scattering models for Sgr A*.

Figure 2.8 shows our detected correlated flux density as a function of baseline length. The ALMA-GBT baseline, probing scales along the minor axis of the source, measures significantly higher correlated flux density than predicted from the Gaussian curves from model fitting to shorter baseline data, shown as the dark blue dashed curves in Figure 2.8 [O16; B19]. This enhancement could either indicate non-Gaussian intrinsic structure (e.g., a compact core with a diffuse halo) or a non-Gaussian scattering kernel (requiring an inner scale $r_{in} \ll 5 \times 10^9 \lambda \approx 5000$ km). For example, the ALMA-GBT measurements are comparable to the values predicted for an anisotropic Gaussian intrinsic source combined with the J18 scattering model, which has $r_{in} = 800$ km, shown as the light blue dotted curves in Figure 2.8.

We also detect correlated flux density on baselines that are expected to entirely resolve the scattered source. Here, the enhanced flux density indicates the presence of image substructure that can either be intrinsic or scattering-induced. For scattering substructure, the signal is expected to be significantly stronger for baselines that are aligned with the major axis of the scattering (see Figure 2.8). The two candidate scattering models presented in Section 2.2.2 (with different spectral index $\alpha$ and inner scale $r_{in}$ governing the refractive noise power spectrum) predict different levels of refractive noise along both the major and minor axes of the scattering: the GS06 model predicts, on average, nearly one order of magnitude more correlated flux density on long baselines than the J18 model. However, our most sensitive detections (ALMA-VLBA/GBT) are along the minor axis of the scattering.

The mean visibility amplitude (after debiasing to account for thermal noise) on baselines longer than 1.8 $\lambda$ is 6 mJy. Because this amplitude may contain contributions from both scattering substructure and intrinsic substructure, it only determines an upper limit on the level of refractive noise from scattering substructure. Moreover, even if there were no intrinsic substructure contribution on these baselines, the 6 mJy signal would still not directly determine the level of refractive noise because refractive noise is stochastic; the inner 95% of visibility amplitudes sam-
Figure 2.8: Noise-debiased correlated flux density for Sgr A* as a function of projected baseline length for data after self-calibrating to the Gaussian source from O16; B19 using only baselines shorter than 0.75 Gλ. Because the a-priori calibration for the GBT was excellent (see Table 2.1), we did not apply the derived GBT gains. Baseline labels are ordered by median baseline length. Intra-European baselines are entirely constrained by the self-calibration and are omitted here for clarity. Dark blue dashed curves show expected visibilities along the major and minor axes for an anisotropic Gaussian source with FWHM of (215 µas, 140 µas); light blue dotted curves show the visibility expected for an anisotropic intrinsic Gaussian source (140 µas, 100 µas) scattered with the non-Gaussian kernel from the J18 scattering model, which has an image size (via 2nd mom.) of (216 µas, 132 µas); red curves show the expected renormalized refractive noise along the major and minor axes for the J18 and GS06 scattering models. Detections on baselines longer than 1Gλ are only obtained for baselines oriented close to the minor axis of the scattering kernel (all are ALMA-VLBA/GBT). Labeled black triangles show upper limits (4σ) on four sensitive baselines at other orientations, all of which have corresponding detections for our calibrators. Colored lines show the anisotropic Gaussian model curves for the corresponding data.
Chapter 2: The Size, Shape, and Scattering of Sgr A*

pled on a single baseline over different scattering realizations will fall in the range \([0.16, 1.9] \times \hat{\sigma}\), where \(\hat{\sigma}\) is the RMS “renormalized” refractive noise [i.e., refractive noise after removing the contributions of flux modulation and image wander, which our observations would absorb into the overall calibration; see J18].

We can tighten the constraints on refractive noise by combining samples from many baselines, although these will be correlated [see Johnson & Narayan, 2016]. Following the Monte Carlo approach of J18, we find that combining baselines longer than \(1.8 \text{ G}\lambda\) gives a 95% confidence range for the mean amplitude of refractive noise on a baseline with \((u, v) = (1.167, -1.638) \times 10^9 \lambda\) of 3-18 mJy if the 6 mJy of correlated flux density is entirely from refractive noise. For comparison, the J18 model predicts a mean refractive noise amplitude of approximately 7 mJy on this baseline, while the GS06 model predicts a mean refractive noise of 60 mJy on this baseline. Thus, the GS06 model is incompatible with our measurements. The GS06 model also significantly over-predicts the signal on our baselines oriented closer to the major axis, for which our measurements only provide upper limits (labeled black triangles in Figure 2.8).

If the minor axis detections are from scattering substructure, then they would represent the first detections of substructure along this axis. The presence of substructure along the minor axis requires that magnetic field variations transverse to the line of sight are not restricted to a narrow angular range (the field wander is more likely to sample all angles, but with a preference for angles that are aligned with the minor axis of the scattering). Minor axis substructure would eliminate, for example, the “boxcar” model for refractive fluctuations in Psaltis et al. [2018], which describes magnetic field wander as a uniform distribution over a limited range of angles.

2.6.2 Constraints on accretion flow and jet models

The intrinsic image of Sgr A* at 3 mm shown in Figure 2.5 allows us to discriminate between the two main classes of models that now must fit the tight source size and morphology constraints derived from both model-fitting (from previous experiments) and our image-domain measurements. We can explore a small subset of GRMHD simulations to assess possible constraints from our observables. Due to our unconstrained estimate of the PA, we opted to compare the major axis size and the asymmetry (axial ratio), which are independent of the PA of the source on the sky.

Figure 2.9 compares the sizes and morphology of 7/3/1.3 mm images from a sample of 3D-GRMHD simulations of either disk or jet dominated emission, at varying viewing angle with respect to the black hole spin axis, with observational constraints from current (Table 2.3) and previous observations of Sgr A* [see Table 4 in J18]. Model images are generated by combining the dynamical model with ray-tracing and radiative transfer using only synchrotron opacities.

Producing a ray-traced image from single-fluid GRMHD simulations requires providing the electron distribution function (hereafter eDF), which is unconstrained in traditional single-fluid GRMHD simulations. Thermal disk models (“Th disk” in Figure 2.9) assume a thermal, Maxwell-Jüttner eDF and a proton-to-electron temperature ratio\(^3\) \(T_p/T_e = 3\) everywhere [motivated by

\(^3\)Standard GRMHD simulations provide only the fluid pressure, which is dominated by the protons. In a perfect fluid, the pressure in a grid zone gives a proton temperature. For strongly sub-Eddington accretion flows with
Inclination [degrees]

Intrinsic major axis FWHM [as]

This work

Inclination [degrees]

Intrinsic asymmetry (maj/min)

This work

Th disk

Th jet

Th+ disk

Th+ jet

1 mm

3 mm

7 mm

1 mm

3 mm

7 mm

Figure 2.9: Intrinsic size and asymmetry (axial ratio) estimates from observations of Sgr A* at 1, 3 and 7 mm vs. theoretical predictions based on 3D GRMHD simulations of black hole accretion flows. Line color encodes the wavelength of observation and the bands are size and asymmetry bounds from model-fitting [J18]. The upper and lower size and asymmetry image-domain bounds from this work are shown as solid magenta lines. Data constraints at 1 mm extend to a lower asymmetry bound of 1.0. Various line types correspond to models with varying prescriptions for electron acceleration and disk/jet dominated flows generated at each wavelength: “th” for a purely thermal disk or jet dominated emission model, “th+κ” for a thermal model with accelerated particles (1% for disk and 10% for jet) in a power-law (κ) distribution [Mościbrodzka et al., 2009, 2014, 2016; Davelaar et al., 2018]. Left: Intrinsic source sizes as a function of the viewing angle. Right: Intrinsic asymmetry (axial ratio) as a function of the viewing angle.
**Figure 2.10:** Intrinsic size and asymmetry (axial ratio) estimates from observations of Sgr A* at 1, 3 and 7 mm vs. theoretical predictions based on 3D GRMHD simulations of black hole accretion flows. Line color encodes the wavelength of observation and the bands are size and asymmetry bounds from model-fitting [J18]. The upper and lower size and asymmetry image-domain bounds from this work are shown as solid magenta lines. Data constraints at 1 mm extend to a lower asymmetry bound of 1.0. Various line types correspond to models with varying prescription for electron heating and black hole spin generated at each wavelength: “H” for the Howes turbulent cascade prescription, “R” for the Rowan magnetic reconnection prescription, “Lo” for a non-spinning black hole, and “Hi” for a black hole with a dimensionless spin of 0.9375 [Howes, 2010; Rowan et al., 2017; Chael et al., 2018a]. **Left:** Intrinsic source sizes as a function of the viewing angle. **Right:** Intrinsic asymmetry (axial ratio) as a function of viewing angle.
2.6 Discussion

results of Mościbrodzka et al., 2009]. Models denoted as “Th jet” have $T_p/T_e = 20$ in the accretion disk and $T_p/T_e = 1$ along the magnetized jet, which allows the jet to outshine the disk at mm-wavelengths (this jet model has been introduced by Mościbrodzka et al. [2014, 2016]). There is a family of models in-between these two extreme cases. In the models denoted as “Th+κ disk” the eDF is hybrid; 1 percent of all electrons are non-thermal, described by a $κ$ eDF. Adding non-thermal electrons to the emission model results in more extended disk images as the non-thermal electrons produce a diffuse “halo” around the synchrotron photosphere. The “halo” contributes to the disk size estimates [Mao et al., 2017]. Finally, the “Th+κ jet” model is a 3D version of the $κ$–jet model introduced by Davelaar et al. [2018] with 10 percent of jet electrons in a $κ$ eDF. In both hybrid models the $κ$ parameter is set to 4 [see Davelaar et al., 2018, for details].

We find that only disks with a hybrid eDF at moderate viewing angles and both jet-models with viewing angles $\lesssim 20^\circ$ are consistent with 1.3 and 3mm sizes and asymmetry constraints. This limit is consistent with the recent low-inclination constraints derived from orbital motions in near-infrared Sgr A* flares by Gravity Collaboration et al. [2018b] observed with the GRAVITY instrument. In the tested models, the dependency of the source sizes as a function of observing wavelength is shallower than the $θ \sim λ$ dependency estimated from multi-wavelength observations of Sgr A* [Figure 13 in J18]. Hence none of the models that satisfy 1.3/3mm source sizes can account for the 7mm source size.

Although GRMHD simulations of black hole accretion are inherently time-variable, causing the size and asymmetry to fluctuate in time, these changes are smaller than 10 percent. We conclude that current models under-predict the observed 7mm emission size, even when accounting for size and asymmetry fluctuations in time. In simulations, the 7mm photons are emitted from larger radii where the accretion flow structure is less certain due to lower grid resolution, the initial conditions (finite size torus with pressure maximum at $r = 24 \text{GM}/c^2$) and boundary conditions of the simulation that only allow for plasma outflows. These issues as well as the electron acceleration should be addressed by future radiative GRMHD simulations of Sgr A*.

We also explored another set of 3D simulations from Chael et al. [2018a], performed with the two-temperature, radiative GRMHD code KORAL [Sądowski et al., 2013, 2014, 2017, see Figure 2.10]. Unlike the simulations presented in Figure 2.9, where the electron temperature (and potential non-thermal component) is assigned to the simulation in post-processing, KORAL evolves the electron temperature throughout the simulation self-consistently with contributions from radiative cooling, Coulomb coupling, and dissipative heating. While the physics of radiation and Coulomb coupling is well understood, the dissipative heating of electrons and ions is governed by unconstrained plasma microphysics that occurs at scales far smaller than the grid scale of the simulation.

Chael et al. [2018a] investigated two different physical prescriptions for the electron dissipative heating. The first prescription is the Landau-damped turbulent cascade model of Howes [2010]. Since this prescription primarily heats electrons in regions where the plasma is highly magnetized, it produces prominent emission from the jet and outflow of the GRMHD simulations at 3.5 mm[see

$L_{\text{bol}}/L_{\text{edd}} \approx 10^{-8}$, protons and electrons are not necessarily well coupled by Coulomb collisions. In these GRMHD simulations the electron temperatures are not self-consistently computed but they are essential in calculating synchrotron emission. The electron temperature is parameterized by a coupling ratio, $T_p/T_e$, between the proton and electron temperature.
The other prescription for electron heating investigated in Chael et al. [2018a] is based on particle-in-cell simulations of particle heating from magnetic reconnection presented in Rowan et al. [2017]. This prescription heats electrons and ions equally and only in highly magnetized regions, resulting in cooler jet regions with less emission than the disk. In total, Chael et al. [2018a] presented four simulations spanning the two heating prescriptions considered (“Howes” or “H” for the turbulent cascade prescription of Howes 2010 and “Rowan” or “R” for the reconnection prescription of Rowan et al. 2017) and two values of the dimensionless black hole spin ($a = 0$ for “Lo”, and $a = 0.9375$ for “Hi”).

Figure 2.10 shows that all four models presented in Chael et al. [2018a] fit the 1.3 mm constraints and mostly fit the 3 mm image-domain constraints. However, only the H-Hi and R-Lo models fit the model-fitting 3 mm range at moderate viewing angles, and all models fail to match 7 mm constraints. However, these simulations were only run over a relatively short time, and inflow equilibrium in the disk was only established up to $\sim 20 \ R_{Sch}$, while the 7 mm emission extends to $\sim 35 \ R_{Sch}$. To accurately compare the predictions from these two heating models with predictions at 7 mm and longer wavelengths, the simulations will have to be run longer using initial conditions adapted to producing an accretion disk in equilibrium past $20 \ R_{Sch}$.

Figure 2.11 demonstrates the plausible range of intrinsic source sizes vs. asymmetries at 3 mm for all of the models we have explored. Here it is evident which models fall into the permitted
Figure 2.12: Scan-averaged closure phases for Sgr A* on three triangles (LA-KP-FD, GBT-KP-FD, ALMA-GBT-FD) with predictions for a thermal+$\kappa$-distribution disk model (th+$\kappa$ disk, $i = 60^\circ$), thermal+$\kappa$-distribution jet model (th+$\kappa$ jet, $i = 90^\circ$) and thermal jet model (th jet, $i = 5^\circ$ and $i = 90^\circ$), where $i$ is the inclination. Each model is shown without scattering (top), with ensemble-average scattering (center), and with a single realization of scattering (bottom). Note that ensemble-average scattering does not affect closure phase, and even a single realization of the scattering has little effect on the closure phases for these triangles.
region. Given that our modeling does not involve any detailed parameter fitting, the agreement between models and observables is encouraging. Disk and jet models with different heating prescriptions are also likely to have distinct polarimetric characteristics that can be compared to observables [e.g., Gold et al., 2017; Mościbrodzka et al., 2017].

Furthermore we can directly compare closure phases from the different models with those presented in Section 2.5. Closure phases observed are an additional robust criterion to discriminate between models: they are independent of imaging assumptions, the beam of the observations, and scattering effects. In Figure 2.12 we compare the scan-averaged closure phases from Sgr A* for the three representative triangles to four example models: the thermal+κ disk model at an inclination of 60° and the thermal jet model at an inclination of 5°, which fit the major axis and asymmetry bounds given by the 3.5 mm observations; and the thermal+κ jet model and thermal-only jet model, both at an inclination of 90°, which do not fit the 3.5 mm bounds. We simulated observations of the four different models with the same stations and coverage as our GMVA+ALMA dataset and compare the closure phases for the original model images (Figure 2.12; top panel), for the model scattered with the J18 refractive scattering (center panel), and for the “ensemble average” models scattered only with the scattering kernel (no refractive noise; bottom panel). We find that for the small and medium triangles it is very difficult to distinguish between models as they all have closure phases near zero, similar to our measurements [Fraga-Encinas et al., 2016]. However, for the large triangle (ALMA-GBT-FD), two models show strong non-zero closure phases: the thermal+κ and the thermal-only jet models at 90° inclination. Interestingly, these are also the example models that do not fit the intrinsic asymmetry and size bounds from 3.5 mm. We also find that interstellar scattering as modeled by J18 does not strongly affect intrinsic closure phase: for both the ensemble average and fully scattered cases, the two jet models at 90° inclination clearly deviate from what is measured on the largest triangle. The Howes and Rowan models are omitted from the comparisons in Figure 2.12 for clarity, as they are all very symmetrical and compact: their closure phase behavior is similar to the thermal+κ disk and the thermal jet models nearly or fully pointed along the line of sight.

While our comparisons to simulations are limited to a handful of GRMHD models, they demonstrate the strong constraints provided by multi-wavelength measurements of size, shape and point-symmetry of Sgr A*.

2.7 Summary

We have presented observations of Sgr A* using ALMA in concert with the GMVA at 86 GHz. These are the first observations to use ALMA as part of a VLBI array, improving the angular resolution for observations of Sgr A* at this frequency by more than a factor of two. The improved resolution and sensitivity have allowed us to reconstruct an intrinsic image of Sgr A* for the first time at this frequency, which is also the first image of Sgr A* for which the scattering is subdominant to intrinsic structure. We find that the intrinsic image of Sgr A* has an asymmetry (axial ratio) of \( 1.2^{+0.3}_{-0.2} \) and a major axis of \( 120 \pm 34 \) μas, although we cannot constrain the position angle because of the highly symmetric intrinsic source.

We have demonstrated that the geometrical properties of the intrinsic image and observed closure phases tightly constrain accretion flow models onto Sgr A*. Our measurements require
models to have symmetrical morphology, 86 GHz radio emission spanning 12 ± 3.4 Schwarzschild radii, and closure phases close to zero on the triangles sampled in our observation. For the eight theoretical simulations we have considered at 3.5 mm, our data are compatible with disk models at all inclinations and jet models fully or nearly pointed along the line of sight. None of the simulations we consider is able to simultaneously match size and asymmetry limits from 1, 3, and 7 mm observations due to the relatively small domain simulated by state of the art 3D GRMHD models.

While GRMHD models are promising to describe emission near the horizon, semi-analytical models for the accretion flow and jet can be more readily extended to larger domains [e.g., Broderick et al., 2016; Gold et al., 2017; Pu & Broderick, 2018]. In addition, it is more straightforward to explore parameter dependencies for semi-analytic models. The model of Broderick et al. [2016] is compatible with our 3.5 mm size and asymmetry estimates. Exploring whether these models can be compatible with the full set of multi-wavelength size and asymmetry constraints for Sgr A* is a promising avenue for continued study.

In addition to the overall image morphology, we have discovered non-Gaussian structure along the minor axis of Sgr A*, hinting at either a non-Gaussian intrinsic source or a non-Gaussian scattering kernel. Comparisons of the observed visibility amplitudes against two scattering models showed that the scattering model presented by Goldreich & Sridhar [2006] over-predicts the correlated flux density on long baselines to ALMA (1.8-2.4 Gλ). This model also overpredicts the flux density on east-west baselines longer than 1 Gλ. Thus, this model for the scattering of Sgr A* is conclusively ruled out by our observations. The exclusion of the GS06 model shows that refractive scattering is likely to weakly affect 1.3 mm images with the EHT.

The scattering model presented by Johnson et al. [2018], on the other hand, predicts comparable levels of refractive noise to the excess flux density we have observed on baselines above 1.8 Gλ. However, using our single observation with ALMA, we cannot conclusively determine whether those detections are entirely due to refractive noise or if they are a combination of intrinsic source structure and scattering substructure. Continued observations of Sgr A* will elucidate these questions, including deeper VLBI observations at 22 and 43 GHz to better estimate the inner scale from the shape of the scatter-broadening kernel (e.g., I. Cho et al. 2021, in prep), and additional GMVA+ALMA observations that will sample different realizations of the scattering screen.

Acknowledgements
This work is supported by the ERC Synergy Grant “BlackHoleCam: Imaging the Event Horizon of Black Holes”, Grant 610058. We thank the National Science Foundation (AST-1126433, AST-1614868, AST-1716536) and the Gordon and Betty Moore Foundation (GBMF-5278) for financial support of this work. This work was supported in part by the Black Hole Initiative at Harvard University, which is supported by a grant from the John Templeton Foundation. K. A. is a Jansky Fellow of the National Radio Astronomy Observatory. M. K. acknowledges the financial support of JSPS KAKENHI grants No. JP18K03656 and JP18H03721. R.-S. L. is supported by the National Youth Thousand Talents Program of China and by the Max-Planck Partner Group. L. L. acknowledges the financial support of DGAPA, UNAM (project IN112417), and CONACyT, México. I. C. acknowledges the financial support of the National Research Foundation of Korea.
(NRF) via a Global PhD Fellowship Grant (NRF-2015H1A2A1033752). This paper makes use of the following ALMA data: ADS/JAO.ALMA2016.1.00413.V. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This research has made use of data obtained with the Global Millimeter VLBI Array (GMVA), which consists of telescopes operated by the (Max-Planck-Institut für Radioastronomie) (MPIfR), IRAM, Onsala, Metsahovi, Yebes, the Korean VLBI Network, the Green Bank Observatory and the Long Baseline Observatory (LBO). The VLBA is an instrument of the LBO, which is a facility of the National Science Foundation operated by Associated Universities, Inc. The data were correlated at the DiFX correlator of the MPIfR in Bonn, Germany. This work is partly based on observations with the 100-m telescope of the MPIfR at Effelsberg. This work made use of the Swinburne University of Technology software correlator [Deller et al., 2011], developed as part of the Australian Major National Research Facilities Programme and operated under licence.
VLBI imaging of black holes via second moment regularization

S. Issaoun, M.D. Johnson, L. Blackburn, M. Moscibrodzka, A. Chael, and H. Falcke

published 2 September 2019

Abstract

The imaging fidelity of the Event Horizon Telescope (EHT) is currently determined by its sparse baseline coverage. In particular, EHT coverage is dominated by long baselines, and is highly sensitive to atmospheric conditions and loss of sites between experiments. The limited short/mid-range baselines especially affect the imaging process, hindering the recovery of more extended features in the image. We present an algorithmic contingency for the absence of well-constrained short baselines in the imaging of compact sources, such as the supermassive black holes observed with the EHT. This technique enforces a specific second moment on the reconstructed image in the form of a size constraint, which corresponds to the curvature of the measured visibility function at zero baseline. The method enables the recovery of information lost in gaps of the baseline coverage on short baselines and enables corrections of any systematic amplitude offsets for the stations giving short-baseline measurements present in the observation. The regularization can use historical source size measurements to constrain the second moment of the reconstructed image to match the observed size. We additionally show that a characteristic size can be derived from available short-baseline measurements, extrapolated from other wavelengths, or estimated without complementary size constraints with parameter searches. We demonstrate the capabilities of this method for both static and movie reconstructions of variable sources.

3.1 Introduction

Very-long-baseline interferometry (VLBI) is a technique able to achieve high angular resolution imaging through the use of widely separated antennas. Unfortunately, as the observing frequency
is increased, the availability of suitable sites on Earth is greatly reduced, leading to sparse arrays with a high angular resolution but a low spatial dynamic range. In particular, a simple inverse Fourier transform of the visibilities measured by an interferometer, or ‘dirty image’, is dominated by artifacts introduced by sparse sampling of the Fourier plane. Short baselines are particularly important in imaging, as they anchor the flux distribution and provide a crucial link between high-resolution small-scale features and the large-scale extent and morphology of the target. The sparser the array, the more challenging it is to reconstruct images from interferometric measurements. Additionally, weather and technical issues at sites that provide short/mid-range baselines can greatly degrade the ability to image a given data set.

Array sparsity and station-based errors can have dramatic effects on reconstructed images. Thus, the imaging process requires further information and assumptions beyond the visibility measurements from the interferometer. The choice of imaging method imposes additional constraints on the reconstructed image. Here, we will focus on extending the method of regularized maximum likelihood (RML) that performs well under sparse sampling conditions and does not involve direct inverse Fourier transforms of the data in the imaging process.

In this paper we present an algorithmic contingency to array sparsity and site issues in the form of a second moment regularization function. That is, the compactness of the source can be expressed as the second moment of the source brightness distribution [Moffet, 1962; Burn & Conway, 1976], which can be constrained to match, for example, confident source size measurements from short baselines of previous experiments or epochs. Enforcing this source size constraint supplements limited short-baseline information while fitting to long-baseline smaller scale structure from newer observations.

The Event Horizon Telescope (EHT), observing at a frequency of 230 GHz [Event Horizon Telescope Collaboration et al., 2019a,b], is a prime example of a high-frequency VLBI imaging experiment with image uncertainties dominated by the effects of sparse coverage. The EHT currently has only a single short/mid-range VLBI baseline, joining the Large Millimeter Telescope Alfonso Serrano (LMT) in Mexico to the Submillimeter Telescope (SMT) in Arizona. Recent observations with the EHT have shown that the LMT is difficult to calibrate, giving baselines with large measurement uncertainties dominated by uncharacterized station behavior in 2017 [Event Horizon Telescope Collaboration et al., 2019c,d].

Although the EHT observes a number of non-horizon-scale sources in conjunction with the Atacama Large Millimeter/submillimeter Array (ALMA), its primary targets are the two supermassive black hole candidates in the Galactic Center, Sagittarius A* (Sgr A*), and at the center of the radio galaxy M87. At the frequency of the EHT, these two sources are very compact, with sizes on the sky historically measured with three stations, in California, Arizona, and Hawaii, in early EHT observations, and are thus ideal imaging targets for second moment regularization [Doeleman et al., 2008; Fish et al., 2011; Doeleman et al., 2012; Akiyama et al., 2015; Johnson et al., 2015; Lu et al., 2018]. Near-zero closure phases on the California–Arizona–Hawaii triangle are indicative of the source compactness and symmetry on scales of a few tens of μas [Akiyama et al., 2015; Fish et al., 2016]. The California–Arizona baseline provided the short-baseline measurements needed to constrain the compactness and size of the sources in the visibility domain. Recent observations of M87 in 2017 also found a source size of $\sim 40$ μas consistent with previous measurements [Event Horizon Telescope Collaboration et al., 2019a,b,c,d,e,f].
3.2 Background

For Sgr A∗, the source size is also well-constrained at lower frequencies due to its compactness and dominant diffractive scattering [Shen et al., 2005; Bower et al., 2006; Lu et al., 2011a; Johnson et al., 2018]. VLBI observations at 86 GHz taken one month apart give fitted Gaussian source sizes for the scattered image of Sgr A∗ with < 10% difference [Ortiz-León et al., 2016; Brinkerink et al., 2019]. At this frequency, while the small scale structure is expected to vary, the large-scale information, dominated by the size of the scattering kernel, should be stable from epoch to epoch [Johnson et al., 2018].

Second moment regularization merges the benefits of model-fitting with the flexibility of imaging: compared to self-calibration to a known model, it does not actually modify the measured visibilities used for the imaging process or enforce a model-dependent solution, but instead provides additional information to improve image quality. The regularization constrains the spread of flux density to a motivated region in the image, discouraging non-physical morphology driven by fits to long-baseline data and accelerating convergence toward a plausible image. It is a natural extension of imaging tools that add source information in the imaging process in RML methods: a total flux constraint is in fact the zeroth moment of the image; an image centroid specification corresponds to the first moment of the image; and a short-baseline source size completes the picture by constraining the image second moment. The implementation of second moment regularization can be done in conjunction with other tools and constraints in RML, for both static and movie reconstructions. Furthermore, as the constraint function acts on the image itself and does not modify the visibility data, it can be used with any choice of data product, including minimally-calibrated closure phases and amplitudes.

The paper is structured as follows. We present the mathematical background to motivate the regularization in Section 3.2. We outline the method, assumptions, and physical motivation in Section 3.3. In Section 3.4 we demonstrate the improvements in image quality and fidelity using the regularization with or without a priori knowledge of the source size. We present possible applications of the second moment regularization to more sophisticated imaging techniques for scattering mitigation and movie reconstructions in Section 3.5. A summary is given in Section 3.6.

3.2 Background

By the van Cittert-Zernike theorem, an interferometer samples complex visibilities corresponding to Fourier components of an image [van Cittert, 1934; Zernike, 1938]. Consequently, $n^{th}$ moments of an image correspond to $n^{th}$ derivatives of the visibility function at the origin. Specifically, an interferometric visibility $V(u)$ on a baseline $u$ can be written as [e.g., Thompson et al., 2017]

$$V(u) = \int d^2 \mathbf{x} I(\mathbf{x}) e^{-2\pi i u \cdot \mathbf{x}},$$

(3.1)

where $I(\mathbf{x})$ is the brightness distribution on the sky, and $\mathbf{x}$ is an angular unit.

From this expression, $V(0) = \int d^2 \mathbf{x} I(\mathbf{x}) \in \mathbb{R}$ gives the total flux density of the image (the 0th moment). Likewise, the phase gradient of the visibility function at zero baseline gives a vector proportional to the centroid of the image,

$$\nabla V(u)|_{u=0} = -2\pi i \int d^2 \mathbf{x} \mathbf{x} I(\mathbf{x})$$

$$= -2\pi i V(0) \mu,$$

(3.2)
where $\mu$ is the image centroid (the normalized 1st moment):

$$
\mu = (\bar{x}, \bar{y}) = \frac{\int d^2x I(x)x}{\int d^2x I(x)}. \tag{3.3}
$$

Because the image is real, the gradient $\nabla V(u)|_{u=0}$ is purely imaginary. For images that are positive (e.g., images in total intensity), the visibility function must take its maximum amplitude at the origin. More generally, the visibility function is Hermitian; thus, its amplitude must always have a vanishing gradient at the origin because of the conjugation symmetry $V(u) = V^*(-u)$.

The second derivative, or Hessian, of the visibility amplitude function at zero baseline gives a matrix (see Appendix 3.A.1):

$$
\nabla \nabla^\top |V(u)|_{u=0} = -4\pi^2 \int d^2x I(x)(x - \mu)(x - \mu)^\top = -4\pi^2V(0)\Sigma, \tag{3.4}
$$

where $\Sigma$ is the normalized second central moment (or covariance matrix) of the image. We show in Appendix 3.A.1 that this expression is equivalent to the curvature of the centered complex visibility function [see also Moffet, 1962; Burn & Conway, 1976]. The visibility amplitude function is a more natural data product to use for observations with non-astrometric VLBI arrays such as the EHT, where there is no absolute phase information due to strong differential atmospheric propagation effects between sources, and thus no directly measured full complex visibilities. Therefore it is useful for us to determine image moments directly from the visibility amplitude function, which is measured.

The image covariance matrix $\Sigma$ can be more intuitively expressed in terms of its principal axes, corresponding to the perpendicular axes about which the second moment reaches its maximum [Hu, 1962]. The matrix has two eigenvalues $\lambda_{\text{min}}$ and $\lambda_{\text{maj}}$, and can be diagonalized as follows:

$$
\Sigma = R_\phi \begin{pmatrix} \lambda_{\text{min}} & 0 \\ 0 & \lambda_{\text{maj}} \end{pmatrix} R_\phi^\top, \tag{3.5}
$$

where $R_\phi$ is the rotation matrix based on the position angle east of north $\phi$ of the major principal axis (Appendix 3.A.2). The eigenvalues of the covariance matrix are the variances of the normalized image projected along the principal (major and minor) axes. The correspondence between $\lambda_{\text{maj}}$, $\lambda_{\text{min}}$, $\phi$ and the individual terms of $\Sigma$ is given in Appendix 3.A.2.

Following Equation 3.1, we can fully express the visibility function as a Taylor expansion in its derivatives. Each $n+1^{\text{th}}$ term of the Taylor expansion is proportional to the $n^{\text{th}}$ moment of the visibility function (see Table 3.1). At zero baseline, only the zeroth moment remains. We choose the coordinate system such that the centroid of the image is at the origin, and the first moment of the visibility function (the second term of the Taylor expansion) vanishes. At short baseline, the centered complex visibility function is therefore dominated by the quadratic term. The Taylor expansion of the visibility function at short baseline becomes:

$$
V(u) \simeq V(0) - 2\pi^2 \int d^2x (u \cdot x)^2 I(x) \\
\simeq V(0) - 2\pi^2V(0)u^\top \Sigma u. \tag{3.6}
$$
Table 3.1: Correspondence of the mass, center of mass and moment of inertia in the image and visibility domains.

<table>
<thead>
<tr>
<th>Physical Analog</th>
<th>Image Domain</th>
<th>Visibility Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>Total Flux $\int I(x)d^2x$</td>
<td>Peak Visibility $V(0)$</td>
</tr>
<tr>
<td>Center of Mass</td>
<td>Centroid $\mu V(0)^{-1} \int xI(x)d^2x$</td>
<td>Phase Gradient $(2\pi iV(0))^{-1} \nabla V(u)</td>
</tr>
<tr>
<td>Moment of Inertia</td>
<td>Covariance $\Sigma V(0)^{-1} \int xx^\top I(x)d^2x$</td>
<td>Amplitude Curvature $(-4\pi^2V(0))^{-1} \nabla \nabla V(u)</td>
</tr>
</tbody>
</table>
Equation 3.6 describes the visibility function behavior on short baselines entirely in terms of the total flux $V(0)$ and the second moment covariance matrix $\Sigma$ projected along the baseline direction. These parameters also describe a unique visibility function of a Gaussian source with total flux $V(0)$, and major/minor axes sizes and orientation prescribed by the same covariance matrix. We show this by comparing the general complex visibility function to that for a Gaussian source. For the simplest case of an isotropic Gaussian source of standard deviation $\sigma$ with the same total flux $V(0)$, we have the following intensity pattern on the sky and corresponding visibility function:

$$I_{\text{gauss}}(x) = \frac{V(0)}{2\pi\sigma^2} e^{-|x|^2/2\sigma^2}, \quad (3.7)$$

$$V_{\text{gauss}}(u) = V(0)e^{-2\pi^2|u|^2\sigma^2}. \quad (3.8)$$

More generally, an anisotropic Gaussian with a covariance matrix $\Sigma$ gives:

$$I_{\text{gauss}}(x) = \frac{V(0)}{2\pi\sqrt{|\Sigma|}} e^{-x^\top\Sigma^{-1}x}, \quad (3.9)$$

$$V_{\text{gauss}}(u) = V(0)e^{-2\pi^2u^\top\Sigma u}. \quad (3.10)$$

Taking the Taylor expansion of the anisotropic Gaussian visibility function at short baselines, the first two terms dominate:

$$V_{\text{gauss}}(u) \simeq V(0) - 2\pi^2 V(0)u^\top\Sigma u. \quad (3.11)$$

We thus obtain an equivalence of the behavior of the general visibility function (Equation 3.6) and the Gaussian visibility function (Equation 3.11) at short baselines. This relation allows us to translate the second moment covariance matrix of the general visibility function to the covariance matrix of an anisotropic Gaussian, which provides a simple parametrization to describe the second moment in terms of the characteristic source extent. The sizes of the major and minor axes $\theta_{\text{maj}}$ and $\theta_{\text{min}}$ are simply the full widths at half-maximum (FWHMs) of the equivalent Gaussian derived from the variances projected along each principal axis:

$$\theta_{\text{maj}} = \sqrt{8 \ln(2) \lambda_{\text{maj}}}, \quad (3.12)$$

$$\theta_{\text{min}} = \sqrt{8 \ln(2) \lambda_{\text{min}}}. \quad (3.13)$$

The equivalence to the Gaussian also gives a natural break-off point where the characteristic source size constraint from the second moment ceases to be a good approximation to the full visibility function: the $1/e$ point determining the resolvability of a Gaussian translates to the baseline length at which the visibility amplitude reaches $V(0)/e$. Baseline lengths longer than the $1/e$ point will lead to higher order terms of the Taylor expansion dominating the behavior and sampling finer structure in the image. We employ the $1/e$ point as a conceptual and visual limit for the source size constraint applied via the second moment regularization. It is not a hard cut-off enforced by the imaging process.

In Figure 3.1, we demonstrate the behavior of the normalized visibility amplitudes sampled along the source major axis as a function of projected baseline length for three images with distinctly different structure but an identical second moment. The behavior on short baselines
3.3 Method

RML focuses on pixelized reconstructions of the image, iteratively maximizing an “objective function”, which is analogous to a log posterior probability function. This function is a weighted (via “hyperparameters”) sum of both $\chi^2_D$ goodness-of-fit data terms, and regularization functions $S_R$, or “regularizers”, governing specific image properties. In this paper, we use the RML method implemented in the eht-imaging Python library [Chael et al., 2016, 2018b], where the objective
function $J(I)$ is minimized via gradient descent, and can be written as:

$$J(I) = \sum_{\text{data terms}} \alpha D \chi^2_D (I) - \sum_{\text{regularizers}} \beta_R S_R (I),$$

(3.14)

where $\alpha_D$ and $\beta_R$ are the input hyperparameters.

Using only five input parameters to the regularization ($V(0)$, $\theta_{\text{maj}}$, $\theta_{\text{min}}$, $\phi$ and $\beta_R$) we can now constrain the second moment of the reconstructed image to match the size constraint provided by the user for RML imaging. In Section 3.3.1 we present our implementation of the second moment regularization function within the 	exttt{eht-imaging} library minimization framework. In Section 3.3.2 we describe the assumptions and physical motivation for second moment regularization using historical observational measurements, known source properties and theoretical expectations.

### 3.3.1 Second moment regularization

Regularization functions in imaging enforce constraints on particular properties of the image, such as image entropy [e.g., Narayan & Nityananda, 1986], smoothness [Bouman et al., 2016; Chael et al., 2016; Kuramochi et al., 2018] and/or sparsity [Wiaux et al., 2009a,b; Honma et al., 2014; Akiyama et al., 2017b,a]. Simple constraints, such as image positivity, image total flux (zeroth moment) or image centering (first moment), are often applied to the image, utilizing known information on the behavior of the total intensity distribution of the source imaged. The implementation of a second moment regularization, constraining the size of the source, is thus a natural extension of common imaging tools that add source information to the imaging process.

We define a regularization function that is minimized when the covariance matrix of the reconstructed image $\Sigma$ matches a user-specified covariance matrix $\Sigma'$. In practice, this latter matrix is computed using user-specified principal axes FWHMs and position angle. We utilize the Frobenius norm to determine a penalty function that quantifies the difference between the user-specified and reconstructed covariance matrices:

$$R_{\Sigma} \equiv \text{Tr} \left[ (\Sigma - \Sigma')^T (\Sigma - \Sigma') \right]$$

(3.15)

This regularizer is, by definition, simply the minimization of the difference between two covariance matrices. The procedure for the regularizer implementation in the 	exttt{eht-imaging} library via gradient descent is presented in Appendix 3.B.

### 3.3.2 Assumptions

The second moment regularization operates under a few key assumptions on the properties of the source observed. The main assumption of this method is the compactness of the source. In order to get a quadratic fall-off in the visibility function, as shown in Section 3.2, the source must be compact and resolved on longer baselines of the interferometer. This method would break down for point sources or sources with complex morphology and diffuse flux on large scales.

Another assumption concerns the stability of the source size across multiple epochs. The input axis sizes and position angle for the regularization will only be valid if the source does not vary significantly in size between observations. The source size input is typically derived from observations where weather conditions, coverage, and station performance on short baselines...
were adequate for higher precision model fitting. The source size can then be used for data sets with larger uncertainties to improve the fidelity and convergence of the imaging process. This assumption is well-motivated for the compact sources observed with the EHT:

- Sgr A* at 86 GHz, has been model-fitted with varying precision over two decades, with little variation in the obtained source size parameters, [Rogers et al., 1994; Krichbaum et al., 1998; Doeleman et al., 2001; Shen et al., 2005; Lu et al., 2011a; Ortiz-León et al., 2016; Brinkerink et al., 2019]

- Sgr A* at 230 GHz has been measured to be compact and stable in size between 2007 and 2013 [Doeleman et al., 2008; Lu et al., 2018; Johnson et al., 2018],

- M87 at 230 GHz has been measured to be compact and stable in size over a decade [Doeleman et al., 2012; Akiyama et al., 2015; Event Horizon Telescope Collaboration et al., 2019a,b,c,d,e,f].

It is worth noting that this assumption breaks down for sources with multiple bright components moving relative to each other, as is common for multi-epoch images of bright jet sources. An overall size measurement from a single epoch would not translate to other observations due to components appearing or moving outward, changing the source morphology significantly between observations. The quadratic fall-off approximation until the \( \frac{1}{e} \) point would also not hold for two separated point sources, which do show a quadratic fall-off in the visibility amplitudes but the amplitudes would quickly evolve to more complex structure on longer baselines that could be identified as the behavior of two point sources interfering. The method is most effective whenever the emission is confined within a single compact region or on multiple scales that are substantially separated, and particularly if the scale of the emission in the image is comparable to the resolution of the array.

We also assume that the extent of the source does not significantly vary within a single epoch. For static imaging of slow-varying sources, it suffices to assume that the average size of the source matches the input, but this has further implications on reconstructions of variable sources within a single epoch. The structural variability on short timescales should be contained within the region constrained by the second moment. This is an issue particularly for imaging Sgr A*, as the source is known to vary on timescales of minutes, much shorter than the length of a single observing epoch. We assess the degree of variability of the source extent in quiescent (non-flaring) models of Sgr A* using general relativistic magnetohydrodynamic (GRMHD) simulations of variable emission on horizon scales [Figure 3.2; Mościbrodzka & Gammie, 2018]. In Figure 3.3, we show the variation in the principal axes FWHMs for a typical GRMHD simulation of the accretion flow of Sgr A* at 230 GHz, both excluding and including the effects of scattering due to the interstellar medium in our line of sight [Johnson, 2016; Johnson et al., 2018]. Although the simulation shows structural changes in the source morphology, deviations about the mean FWHM remain below 10% for both the model and scattered simulation principal axes.

Furthermore, the emitting gas around supermassive black holes in low-luminosity active galactic nuclei becomes optically thin as we increase the observing frequency. The source extent is therefore dominated by the black hole shadow and Doppler-boosted features at higher frequencies [Falcke et al., 2000]. This behavior is shown in Figure 3.4 for the GRMHD simulation of the
Figure 3.2: Left: 230 GHz GRMHD simulation of Sgr A* [Mościbrodzka & Gammie, 2018]. Right: Same simulation including the effects of interstellar scattering [Johnson, 2016; Johnson et al., 2018].

Figure 3.3: Principal axes FWHMs as a function of time for the simulation of Sgr A* in Figure 3.2 [Mościbrodzka & Gammie, 2018]. The solid lines show sizes for the simulation, the dotted lines show sizes for the simulation including the effects of interstellar scattering [Johnson, 2016; Johnson et al., 2018]. The scattering major axis is aligned with the source minor axis, and thus the scattering kernel slightly dominates the minor axis size, which stabilizes the minor axis FWHM time series. The sizes were obtained from measurements of the image second moment per frame. For all four size trends, the deviation about the mean size is < 10%.
3.4 Demonstration

The second moment regularization can be used with informed size constraints from previous experiments, GRMHD simulations, or achromatic features from other observing frequencies. In this section, we demonstrate how the second moment regularization adds information to the imaging process if the data set to be imaged has no short baselines. For all following tests, we use a high $\beta_R = 10^5$, such that the input source size is strongly constrained in the imaging process. To put this value into perspective, $\beta_R = 10^5$ would cause a $\sim 10\%$ difference between the input and reconstructed source sizes to be penalized equivalently to a change in reduced $\chi^2$ of $\sim 1$ in our imaging procedure. This regularization weight tends to drive the second moment of reconstructed images to be within 20% of the input values, therefore allowing some flexibility for the imaging process to deviate from the input second moment toward morphology favored by the available data.

Figure 3.4: Geometric mean FWHM of principal axes as a function of frequency for the ray-traced simulation of Sgr A* in Figure 3.2 [Mościbrodzka & Gammie, 2018]. The blue curve shows size evolution for the simulation, the red curve shows size evolution for the simulation including the effects of interstellar scattering [Johnson, 2016; Johnson et al., 2018]. The sizes were obtained from measurements of the image second moment per frequency bin of 20 GHz. The change in size with increasing frequency becomes greatly reduced at frequencies above 300 GHz, where the size of the source is dominated by the achromatic black hole shadow and the Doppler boosted features [Falcke et al., 2000].

quiescent accretion flow of Sgr A* observed at frequencies from 80 GHz to 1 THz [Mościbrodzka & Gammie, 2018]. At frequencies of $\sim$300 GHz and above, the source size changes very little with increasing frequency. These achromatic properties motivate the extrapolation of a source size from lower-frequency observations with short baselines, such as the EHT at 230 GHz, to higher-frequency imaging experiments such as the upcoming EHT at 345 GHz [Event Horizon Telescope Collaboration et al., 2019b; Doeleman et al., 2019].
Chapter 3: VLBI imaging of black holes via second moment regularization

Figure 3.5: $(u, v)$ coverage for simulated observations of Sgr A* with the EHT 2017 array at 230 GHz. The magenta disk represents the range of $(u, v)$ constrained by the second moment regularization, with the boundary at the $1/e$ point of the corresponding visibility amplitude function for Sgr A* assuming an isotropic source of $60 \mu$as FWHM [Johnson et al., 2018].

Figure 3.6: Visibility amplitudes for a model image of a semi-analytic advection-dominated accretion flow (ADAF) model of Sgr A* [Broderick et al., 2011] with a FWHM of $\sim 60 \mu$as as a function of $(u, v)$ distance sampled by the EHT in 2017 with and without the LMT (affecting mid-range baselines). The regularizer $R_\Sigma$ governs the visibility amplitude behavior at short baselines until the $1/e$ point. This allows us to constrain and correct limitations and uncertainties in LMT calibration based on the expected behavior of the LMT–SMT mid-range baseline.
3.4 Demonstration

In Section 3.4.1 we show improvements to the reconstructions when the source size is known. In Section 3.4.2 we study the image quality and fidelity dependence on the assumed size in the regularization. Finally in Section 3.4.3 we demonstrate that high fidelity images can be obtained without a priori knowledge of the source extent via input parameter searches.

3.4.1 Imaging with complementary size constraints

In Figure 3.5, we illustrate the domain in which the second moment regularization \( R_\Sigma \) operates. The \((u, v)\) coverage is that of a typical observation of Sgr A* with the EHT at 230 GHz. Assuming a source extent of 60 \( \mu as \) from previous observations [Johnson et al., 2018], the \(1/e\) boundary of the visibility function for a source with that characteristic size is shown as a disk on the \((u, v)\) coverage. The only EHT baselines that lie within the \( R_\Sigma \) disk are intra-site baselines and the LMT–SMT short VLBI baseline. A single short VLBI baseline is very limited in constraining the overall extent of the source even assuming optimal performance of the telescopes.

We selected a ray-traced image of a semi-analytic advection-dominated accretion flow (ADAF) model of Sgr A* [Broderick et al., 2011] with a similar characteristic size to the Sgr A* observations to assess the performance of the regularizer and to test the robustness of the imaging process as a function of the input parameters \( \theta_{\text{maj}}, \theta_{\text{min}}, \) and \( \phi \). We sample the image with EHT 2017 coverage (Figure 3.5), where we have total flux density estimates from intra-site baselines and a valuable mid-range baseline (SMT–LMT) describing the extent of the source on the sky, as shown in Figure 3.6. We chose to discard all LMT baselines to limit the coverage and remove the constraining mid-range baseline for the regularizer tests. The extent of the source will then solely be enforced by the user-defined \( \theta_{\text{maj}}, \theta_{\text{min}}, \) and \( \phi \) input parameters for \( R_\Sigma \) in the imaging process. It should be noted that imaging without the LMT not only removes short-baseline information on source extent but also long-baseline information on finer features, creating further differences in reconstructed images. The LMT, due to its size and central location, holds a strong weight in triggering decisions, while the SMT is a smaller and well-exercised station and is fairly flexible to various observing conditions. The choice to discard the LMT is thus mainly motivated by the known difficulties, to date, for the station to observe in a wide range of observing conditions and obtain adequate calibration information [Event Horizon Telescope Collaboration et al., 2019c,d]. Removing the SMT instead, for the purposes of these tests, would give similar results due to the lack of short-baseline information.

In Figure 3.7, we show the model crescent image in the left panel, and example reconstructions for four different scenarios in the right panel. The first scenario is a reconstruction of the full EHT observations of the crescent, using closure quantities and visibility amplitudes, and maximizing simple image entropy. In that case, we obtain a good fit to the visibility amplitudes, and we recover an image very similar to the model image. Then, we reconstruct the same observations constraining the image to match the true second moment, as measured on the true image. With this method, we obtain a marginally improved fit to the amplitudes, but visibly less diffuse flux outside the crescent due to the constraint of \( R_\Sigma \). Once we remove the LMT however, the simple imaging with maximum entropy is not able to reconstruct the morphology of the source, although some compact features are reconstructed that enable a decent fit to the visibility amplitudes. When adding \( R_\Sigma \) to the process, the second moment constraint is able to offset the absence of
Figure 3.7: Left: Model image of a semi-analytic ADAF model of Sgr A* [Broderick et al., 2011], contours of 25, 50, and 75% of the peak flux density are shown in white. Right: Tests of the second moment regularizer using the true image parameters as input ($\theta_{maj} = 58 \mu$as, $\theta_{min} = 52 \mu$as, $\phi = 177^\circ$ as measured directly from the model image), $\chi^2$ values are calculated for the data set without the LMT. We additionally give the resulting source size parameters for each reconstruction. Imaging of the example data set with full EHT 2017 coverage shows little difference between the imaging process with and without the second moment regularizer. When the LMT is removed, and thus the mid-range baseline no longer constrains the source size, $\mathcal{R}_\Sigma$ greatly improves the imaging. It should be noted that differences in finer features imaged with and without LMT are expected due to the loss of some long-baseline information from the removal of the LMT.
short baselines and reconstructs an image of improved quality in terms of both image morphology and goodness-of-fit to the amplitudes. This demonstration shows that $R_{\Sigma}$ successfully adds additional information to reconstruct a more physically plausible image even when mid-range baselines are lacking in the underlying data set. The improvement in the amplitude $\chi^2$ also shows that $R_{\Sigma}$ is a useful tool to aid convergence in imaging.

3.4.2 Dependence of reconstructed images on assumed size

In the demonstration of $R_{\Sigma}$ we constrained the second moment to the true size of the source, to enable an accurate reconstruction of the image. However, in practice, the true size of the source is unknown, and is instead approximated from Gaussian model fitting to closure quantities and/or short-baseline visibility amplitudes and extrapolated from historical measurements. We therefore investigate the robustness of the image reconstructions when the input Gaussian parameters are strongly enforced in the imaging process, corresponding to a strong weight of the $R_{\Sigma}$ hyperparameter, while changing input principal axes FWHMs. We demonstrate this dependence by imaging the data set of the crescent model sampled by the EHT 2017 coverage without the LMT, such that the extent of the source is only enforced by the varying inputs to $R_{\Sigma}$. For simplicity, we use a single common imaging script varying only the input principal axes FWHMs. We assume an isotropic source size such that $\theta_{maj} = \theta_{min}$ and $\phi = 0^\circ$, and a range of input FWHMs of $5 - 90 \mu as$.

We utilize two metrics to compare the quality of the reconstructed image to the true model image. The normalized root-mean-square error (NRMSE) of each image is given by:

$$\text{NRMSE} = \sqrt{\frac{\sum_k (I_k - I'_k)^2}{\sum_k I_k^2}},$$

where $I'$ is the intensity of the reconstructed image and $I$ is that of the true image [e.g., Chae et al., 2018b]. If the reconstructed image is identical to the true image, the NRMSE is zero. Therefore, the input FWHM for the reconstruction resulting in the minimum NRMSE in comparison to the true image gives the best fit.

The normalized cross-correlation (NXCORR) is a sliding inner-product of two normalized functions. For fast numerical computation, we determine the cross-correlation of the Fourier transforms of the normalized intensity patterns of the true image $I_{\text{norm}}$ and the reconstructed image $I'_{\text{norm}}$ at different relative shifts $\delta$ across the extent of the images. For each pixel $i$ in the image, we normalize the intensity via:

$$I_{\text{norm},i} = \frac{I_i - \mu_I}{\sigma_I},$$

where $\mu_I$ and $\sigma_I$ are the mean and standard deviation of the intensity distribution in the image. The cross-correlation for a given shift $\delta$ is then given by:

$$\text{NXCORR}(\delta) = |\mathcal{F}^{-1}\{\mathcal{F}\{I'_{\text{norm}}(x)\} \cdot \mathcal{F}\{I_{\text{norm}}(x + \delta)\}\}|.$$

The shift at which the cross-correlation is maximized is then used to output the final NXCORR value for the two images. This method is less sensitive to individual features in the reconstructed
Figure 3.8: Quality of the images obtained with different input FWHM (major and minor axes equal, position angle is zero). The image quality is measured in three ways: (1) the normalized cross-correlation against the true image, or NXCORR; (2) the normalized root-mean-square error against the true image, or NRMSE, shown in the top panel; and (3) reduced $\chi^2$ goodness-of-fits to the three data products used in the reconstructions (visibility amplitudes, closure amplitudes and phases) shown in the bottom panel. NRMSE is more sensitive to subtle differences in the images than NXCORR due to the higher weight associated with large pixel-by-pixel errors and is minimized in a comparable range of input FWHMs to the reduced data $\chi^2$. The narrow range of FWHMs encompasses the true mean source FWHM (magenta vertical line).
Figure 3.9: Cross-comparisons of reconstructed images with varying isotropic input FWHMs using symmetrically normalized root-mean-square error (SNRMSE). The SNRMSE grid shows a greater correspondence of images with input FWHMs near the true mean FWHM of 55 µas, marked by the dashed black lines. The reconstructed images with varying input size (5–90 µas) are all compared to each other, where image 1 and image 2 are the two images to be compared ($I'_1$ and $I'_2$ respectively in Equation 3.19). The diagonal is each image compared to itself. The SNRMSE grid gives a range of plausible input FWHMs for $R_\Sigma$ that result in high fidelity images even when the true source size is unknown.

This behavior is caused by $R_\Sigma$ rapidly reducing the favored set of images to only those that constrain flux within a given region. The region limits that best represent the flux distribution in the true image allow the minimizing process to focus more quickly on the data terms and achieve better reduced $\chi^2$ values within the given imaging conditions. This property also allows us to survey the response of the imaging process and goodness-of-fits to the available data via parameter searches over different favored second moments (and thus favored flux regions) and determine optimal parameters that best represent properties of the data set.
3.4.3 Imaging without complementary size constraints

The NRMSE metric proves to be more sensitive to differences in the image structure than NX-CORR, as shown in Figure 3.8, due to the higher weight associated with large errors in the computation of the NRMSE. For that reason, we have selected NRMSE to score comparisons between the reconstructed images themselves. For this test, we assume that the true image and true FWHM are unknown, as is the case for real experiments. We instead focus on the morphological characteristics that appear in the images based on the underlying data, and how the inputs to $R_\Sigma$ affect the correspondence between reconstructed images. We restructure the metric into a symmetrically-normalized root-mean-square error [SNRMSE; Hanna et al., 1985; Mentaschi et al., 2013] to render the NRMSE independent of the input and comparison image choice:

$$\text{SNRMSE} = \sqrt{\frac{\sum_k (I'_{1,i} - I'_{2,i})^2}{\sum_k I'_{1,i} I'_{2,i}}}.$$ (3.19)

Here $I'_{1}$ and $I'_{2}$ are the two reconstructed images to be compared. In Figure 3.9, we show an SNRMSE grid comparing each reconstructed image to all others, where the diagonal squares correspond to each image compared with itself. We have marked with dashed lines where the mean FWHM of the true image lies. We find that images with input FWHMs near the true FWHM of the source have a better SNRMSE with each other than all other combinations of images. This test enables the user to find a range of characteristic sizes minimizing SNRMSE via a size parameter search. For compact sources that are distinctly elliptical, a one-dimensional size parameter search is useful to quickly sweep through a wide range of sizes and determine a range of plausible sizes for the source extent. A search within that range, varying parameters in two dimensions ($\theta_{\text{maj}}$, $\theta_{\text{min}}$, and $\phi$), can then be carried out to refine the source size estimate for the imaging process.

We find that the use of the regularizer improves the quality of the resulting image even if the input parameters deviate by 20% from the true values. We also find that the strong use of the regularization, when combined with a size parameter search, is able to converge toward the true FWHM values, even when the true source dimensions are unknown. The use of SNRMSE and $\chi^2$ statistics serve well to score individual images and parameters without a priori knowledge of the source extent.

3.5 Applications

In addition to simple static imaging, second moment regularization can easily be coupled to more sophisticated and complex imaging techniques. In Section 3.5.1 we present an example of the use of second moment regularization for scattering mitigation imaging of Sgr A* at longer wavelengths. In Section 3.5.2 we demonstrate how second moment regularization in individual sparse snapshots improves the quality of dynamical reconstructions of variable sources, such as a movie of an orbiting "hot spot" in Sgr A*'s accretion flow.
3.5 Applications

Figure 3.10: Reconstructions of 22 GHz VLBA+GBT observations and their resulting source extents. MK and SC have no detections, and HN and NL are flagged due to their very low sensitivity in this experiment. **Left**: a simple reconstruction of the scattered image without $\mathcal{R}_\Sigma$. **Center**: a reconstruction of the scattered image via stochastic optics [Johnson, 2016], using the scattering model by Johnson et al. 2018. **Right**: a reconstruction with stochastic optics, using $\mathcal{R}_\Sigma$ and the input source size as determined by Johnson et al. 2018 from high-precision Gaussian model fitting: $\theta_{\text{maj}} = 2296 \pm 61 \mu\text{as}$, $\theta_{\text{min}} = 1802 \pm 39 \mu\text{as}$, $\phi = 79.9 \pm 0.2^\circ$. The reconstruction with $\mathcal{R}_\Sigma$ helps constrain the extent of the source in the north–south direction, where measurements are lacking due to the predominantly east–west configuration of the VLBA+GBT.

3.5.1 Scattering mitigation

The second moment constraint in imaging can both be used for data sets where short baselines are lacking, as demonstrated in Section 3.4, and for data sets where short-baseline measurements have large uncertainties due to difficult observing conditions. An example of the latter case is presented in Issaoun et al. [2019b], where observations of Sgr A* at 86 GHz with the Global Millimeter VLBI Array and ALMA (project code MB007) yielded high signal-to-noise (SNR) detections on long baselines but bad weather at select Very Long Baseline Array (VLBA) stations led to poorly constrained short-baseline measurements. Imaging of the source with RML would not have been feasible with these measurements alone, as the large uncertainties in the short-baseline measurements caused flux to spread nonphysically across the reconstructed images. Since the size of Sgr A* on the sky is well studied and known to be affected by anisotropic scatter-broadening from the interstellar medium [Davies et al., 1976; van Langevelde et al., 1992; Frail et al., 1994; Bower et al., 2004; Shen et al., 2005; Bower et al., 2006; Psaltis et al., 2018; Johnson et al., 2018], previous size measurements [Ortiz-León et al., 2016; Brinkerink et al., 2019] were used to constrain the extent of Sgr A* in the imaging process with $\mathcal{R}_\Sigma$. In this manner, we obtained an image that was able to fit new long-baseline detections to ALMA, likely refractive noise from scattering substructure.

The second moment regularization was also implemented in the scattering mitigation code **stochastic optics** developed by Johnson [2016]. Stochastic optics aims to mitigate the effects of scattering to derive an intrinsic (unscattered) image of the source. The code solves for the unscattered image by separating and mitigating the two main components of the Sgr A* scattering screen: the diffractive scattering that causes the image to become a convolution of the true image
and the scattering kernel; and the refractive scattering that introduces stochastic ripples that
further distort the image. The stochastic optics framework therefore simultaneously solves for the
unscattered image and the scattering screen assuming a given model for the diffractive blurring
kernel and the time-averaged refractive properties. The model assumed here is the Johnson et al.
[2018] scattering model, the best-fitting model to Sgr A* observations to date [Issaoun et al.,
2019b].

The implementation of \( R_\Sigma \) in stochastic optics only constrains the size of the scattered source
(Sgr A* as we see it on the sky) based on historical measurements from model fitting, such that the
technique can more accurately mitigate the effects of interstellar scattering to obtain a physically
motivated intrinsic image of the accretion flow of Sgr A* [for further details, see Issaoun et al.,
2019b]. The intrinsic image itself is not directly constrained by the second moment regularization,
but is derived from the combination of the constrained scattered image and knowledge of the
interstellar scattering.

Here we illustrate the use of \( R_\Sigma \) within stochastic optics using a lower frequency data set.
Observations of Sgr A* at 22 GHz with the VLBA+GBT (project code BG221A) showed clear
long-baseline detections of refractive noise from interstellar scattering [Gwinn et al., 2014; John-
son et al., 2018]. These long-baseline detections should translate to substructure in the image,
distorting the intensity pattern seen for Sgr A* away from the scatter-broadened smooth elongated
Gaussian-like morphology. While the scattering substructure is very apparent in the data set, it is
a non-trivial task to successfully show its effects on the image itself and obtain an intrinsic image
of the source. This is due to the imaging process being driven predominantly by the abundance
of intra-VLBA short-baseline measurements in comparison to the few VLBA–GBT long-baseline
detections. We therefore test the addition of \( R_\Sigma \) on this data set, using the source dimensions in
Table 1 of Johnson et al. [2018] from elliptical Gaussian model fitting.

In Figure 3.10, we show three separate reconstructions of the 22 GHz data set. A standard
RML reconstruction of the data set (Figure 3.10 left panel) shows some distortions in the scattered
image, but the morphology remains fairly smooth and elongated. Standard RML imaging cannot
solve for the scattering properties, therefore the procedure is solely focused on obtaining the
highest fidelity scattered image possible from the data set. We will thus treat this image as
our comparison image for this data set. When using stochastic optics however, the imaging
process is more complex, as it is simultaneously imaging the scattered source and solving for
the scattering properties to disentangle scattering from intrinsic source structure. This process
derives a scattered image that is not well-constrained in the north–south direction due to the
configuration of the VLBA+GBT, resulting in a large source image that is not fully converged
to the image obtained from standard RML (Figure 3.10 center panel). Since the scattered image
does not match our expectations of the physical morphology of the source, the derived intrinsic
image should also not be trusted. The challenge is then to improve the convergence of the
imaging component of stochastic optics to quickly obtain a physically motivated scattered image
and therefore undergo a higher-fidelity separation of the scattering and intrinsic structure. When
using \( R_\Sigma \), where the scattered image is constrained to remain within the size obtained by Johnson
et al. [2018] using elliptical Gaussian model fitting, the resulting scattered image is more elongated
in the east–west direction (Figure 3.10 right panel) and showing distortions similar to those of
the standard RML reconstruction. This shows that the use of \( R_\Sigma \) helps the convergence of the
scattered image through stochastic optics to a more physically motivated reconstruction, and thus will give a more realistic underlying unscattered image of the source.

**Figure 3.11:** Reconstruction of a simulated flare using dynamical imaging [Johnson et al., 2017]. From top to bottom: simulated images of a flare with a period of 27 minutes (model B of Doeleman et al. 2009b); simple dynamical imaging without the LMT (no short-baseline points constraining the source extent); dynamical imaging using $R_\Sigma$ without the LMT (the second moment regularization offsets the lack of short-baseline constraints); simple dynamical imaging with full EHT2017 sampling; dynamical imaging using $R_\Sigma$ with EHT2017 sampling. Using $R_\Sigma$ significantly improved the quality of dynamical reconstructions both with the full array and without the LMT. NXCORR values against the model images are shown in the top left corner for each reconstructed snapshot. The variations in the resulting FWHMs of the reconstructed images are visually evident.
3.5.2 Dynamical imaging

There are additional applications for the second moment regularization in movie reconstructions of variable sources where single snapshots have very sparse coverage. We can test the robustness of movie reconstructions with the loss of short baselines using a simulated flare [model B of Doeleman et al., 2009b] with an orbiting period of 27 minutes around the same crescent model as in Section 3.4. We reconstruct movies of the orbiting “hot spot” using dynamical imaging, enforcing temporal continuity between individual frames [for further details, see Johnson et al., 2017]. We reconstruct a movie of the orbiting hot spot for four different scenarios: (1) we use the EHT 2017 array without the LMT, no short baselines are present in the individual snapshots to constrain the source extent; (2) we use the data set without the LMT, but constrain the extent of the source (the dimensions of the crescent model) with $R_\Sigma$, (3) we use the full EHT 2017 to reconstruct the orbit; (4) we use the full EHT 2017 and $R_\Sigma$ to reconstruct the orbit. In Figure 3.11, we show individual frames of the true simulation and of the reconstructed movies for the four different scenarios. The reconstructions without $R_\Sigma$ either yield unphysical source structure dominated by the dirty image (due to the lack of information without LMT) or contain imaging artifacts from flux spreading due to the sparse coverage of individual snapshots. In particular, even with the full EHT 2017 array, dynamical imaging without $R_\Sigma$ shows north–east and south–west artifacts from the dirty image that persist due to the sparse snapshot coverage. The reconstructions with $R_\Sigma$, even without the LMT, are significantly cleaner and more accurately reconstruct the motion and morphology of the simulation, as shown by NXCORR results when compared to the truth simulated images.

3.6 Summary

In summary, we have developed a regularization function $R_\Sigma$, for use in a regularized maximum likelihood framework for interferometric imaging, that constrains the spread of flux in reconstructed images to match input parameters defined by the user. The second moment regularization is a natural extension of common imaging tools, such as image total flux and image centroid constraints (zeroth and first moment respectively), that help to mitigate the missing information problem in high frequency VLBI. The regularization assumes that the source is compact, with a stable size, and is resolved on longer baselines of the interferometer. The validity of these assumptions for the EHT’s primary targets, Sgr A* and M87, are well-motivated by state-of-the-art GRMHD simulations and long-term observational studies. For well-studied sources, this method allows for contingency against weather, a major deterrent for high frequency VLBI, and gives more flexibility for triggering decisions if key short baselines yield poorly constrained measurements or become unavailable during or between observations.

We have shown that $R_\Sigma$ successfully informs the source behavior on short baselines and is defined only by three Gaussian parameters and the regularization hyperparameter. Imaging with $R_\Sigma$ is able to reconstruct high fidelity images fitting to the data products even if the input source dimensions deviate from the true values by up to 20%. The regularization therefore gives a larger flexibility than needed to account for changes in size from, for example, GRMHD simulations of highly variable sources such as Sgr A*. We have also shown that parameter searches over a range
3.A Properties of the visibility function

Non-astrometric VLBI experiments such as the EHT measure visibility amplitudes directly but do not provide absolute phase information. Nevertheless, the zeroth and second image moments are determined from visibility amplitudes alone [i.e., they do not depend on the measured phase; Moffet, 1962; Burn & Conway, 1976]. For instance, the total flux density \( \int I(\mathbf{x})d^2\mathbf{x} = V(\mathbf{0}) = |V(\mathbf{0})| \) because the zero-baseline visibility is real and positive, and therefore equal to its modulus.

More generally, we can express the visibility function as a Taylor expansion of its derivatives:

\[
V(\mathbf{u}) = \int d^2\mathbf{x} I(\mathbf{x}) \left[ 1 - 2i\pi \mathbf{u} \cdot \mathbf{x} - \frac{(2\pi \mathbf{u} \cdot \mathbf{x})^2}{2} + i\frac{(2\pi \mathbf{u} \cdot \mathbf{x})^3}{6} + \frac{(2\pi \mathbf{u} \cdot \mathbf{x})^4}{24} + \cdots \right].
\]

The visibility amplitude function is image-translation invariant. To obtain a Taylor expansion for visibility amplitudes, we choose the image centroid to be at the origin. The first derivative of the
visibility function (thus the second term of the Taylor expansion) then vanishes, giving

\[ V(u) \simeq \int d^2x \ I(x) \left[ 1 - \frac{(2\pi u \cdot x)^2}{2} \right] \]

\[ \simeq V(0) - 2\pi^2 \int d^2x \ (u \cdot x)^2 I(x). \]  \hfill (3.21)

On short baselines (i.e., those with \( u \cdot x \ll 1 \)), the visibility function is then positive and real, so \(|V(u)| \simeq V(u)\). Since \( u = \begin{pmatrix} u \\ v \end{pmatrix} \) and \( x = \begin{pmatrix} x \\ y \end{pmatrix} \), we can expand the inner product of the two vectors:

\[
(u \cdot x)^2 = u^2 x^2 + v^2 y^2 + 2uvxy
\]

\[ = \begin{pmatrix} u \\ v \end{pmatrix}^2 \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \]  \hfill (3.22)

Combining these results with the definition of the covariance matrix \( \Sigma \) (see Appendix 3.A.2), we obtain:

\[ |V(u)| \simeq V(0) - 2\pi^2 \int d^2x \ (u \cdot x)^2 I(x) \]

\[ \simeq V(0) - 2\pi^2 \begin{pmatrix} u \\ v \end{pmatrix}^2 \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \]

\[ \simeq V(0) - 2\pi^2 V(0)u^\top \Sigma u. \]  \hfill (3.23)

The downward curvature of the amplitude function at zero baseline is thus related to the image covariance by:

\[ \nabla \nabla \nabla |V(u)| \big|_{u=0} = \nabla \nabla \nabla V(u) \big|_{u=0} = -4\pi^2 V(0)\Sigma. \]  \hfill (3.24)

### 3.A.2 Image principal axes and visibility curvature

From Equation 3.24, the curvature of the visibility function on short baselines is proportional to the second central moment of the image projected along the baseline direction. The second central moment of the image is naturally expressed as a covariance matrix:

\[ \Sigma = \frac{\int d^2x I(x)(x-\mu)(x-\mu)^\top}{\int d^2x I(x)} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \]  \hfill (3.25)

\[ \Sigma_{xx} = \frac{\int d^2x I(x)(x-x_\mu)(x-x_\mu)}{\int d^2x I(x)}, \]

\[ \Sigma_{yy} = \frac{\int d^2x I(x)(y-y_\mu)(y-y_\mu)}{\int d^2x I(x)}, \]

\[ \Sigma_{xy} = \frac{\int d^2x I(x)(x-x_\mu)(y-y_\mu)}{\int d^2x I(x)} = \Sigma_{yx}. \]

To put the covariance matrix in a more intuitive form, we express it in terms of its principal axes. The image covariance matrix has two eigenvalues, and can be diagonalized as follows:

\[ \Sigma = R_\phi \begin{pmatrix} \lambda_{\text{min}} & 0 \\ 0 & \lambda_{\text{maj}} \end{pmatrix} R_\phi^\top, \]  \hfill (3.26)
where the rotation matrix $R_\phi$, based on the position angle $\phi$ (East of North) of the major principal axis, is given by:

$$R_\phi = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix}. \hspace{1cm} (3.27)$$

The eigenvalues are derived from the quadratic equation:

$$\lambda_{maj} = \frac{\Sigma_{xx} + \Sigma_{yy}}{2} + \frac{\sqrt{4(\Sigma_{xy})^2 + (\Sigma_{xx} - \Sigma_{yy})^2}}{2},$$

$$\lambda_{min} = \frac{\Sigma_{xx} + \Sigma_{yy}}{2} - \frac{\sqrt{4(\Sigma_{xy})^2 + (\Sigma_{xx} - \Sigma_{yy})^2}}{2}. \hspace{1cm} (3.28, 3.29)$$

We can also express each term of the covariance matrix in terms of the eigenvalues and position angle $\phi$:

$$\Sigma_{xx} = \cos^2(\phi)\lambda_{min} + \sin^2(\phi)\lambda_{maj}, \hspace{1cm} (3.30)$$

$$\Sigma_{yy} = \sin^2(\phi)\lambda_{min} + \cos^2(\phi)\lambda_{maj}, \hspace{1cm} (3.31)$$

$$\Sigma_{xy} = (\lambda_{maj} - \lambda_{min})\cos(\phi)\sin(\phi). \hspace{1cm} (3.32)$$

The eigenvalues of the covariance matrix are the variances along the principal axes (major and minor axes).

### 3.B Implementation via gradient descent

#### 3.B.1 Pixel-based derivation of principal axes

Here we present the computation of the covariance matrix for the pixel-based reconstructions from RML. The centroid of an $n \times n$ pixel-based image is given by the following parameters:

$$\bar{x} = \frac{\sum_k x_i I_i}{\sum_k I_i} \text{ and } \bar{y} = \frac{\sum_k y_i I_i}{\sum_k I_i}, \hspace{1cm} (3.33)$$

where $i$ is the pixel number (from 1 to $k$), $I_i$ is the intensity at that pixel, $x_i$ is the x-position and $y_i$ is the y-position of the pixel in the image. The second moment of the image is given by the covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}, \hspace{1cm} (3.34)$$

where

$$\Sigma_{xx} = \frac{\sum_k (x_i - \bar{x})^2 I_i}{\sum_k I_i}, \hspace{1cm} (3.35)$$

$$\Sigma_{yy} = \frac{\sum_k (y_i - \bar{y})^2 I_i}{\sum_k I_i}, \hspace{1cm} (3.36)$$

$$\Sigma_{xy} = \frac{\sum_k (x_i - \bar{x})(y_i - \bar{y}) I_i}{\sum_k I_i}. \hspace{1cm} (3.37)$$
As in Appendix 3.A.2, the image covariance matrix has two eigenvalues and can be diagonalized to obtain the principal axes FWHMs.

3.B.2 Gradient Descent Implementation

We have defined our regularization function via the Frobenius norm:

\[ R_{\Sigma} = (\Sigma_{xx} - \Sigma_{xx}')^2 + (\Sigma_{yy} - \Sigma_{yy}')^2 + 2(\Sigma_{xy} - \Sigma_{xy}')^2. \]  

(3.39)

Within the framework of the eht-imaging library, the objective function is minimized via gradient descent. Therefore, the regularization functions must also individually be minimized via gradient descent. The gradients for the quantities describing the properties of the image introduced thus far, for a given pixel \( j \), are given below:

\[ \frac{\delta \bar{x}}{\delta I_j} = \frac{x_j \sum_k I_i - \sum_k (x_i I_i)}{(\sum_k I_i)^2} = \frac{x_j - \bar{x}}{(\sum_k I_i)}, \]

(3.40)

\[ \frac{\delta \bar{y}}{\delta I_j} = \frac{y_j \sum_k I_i - \sum_k (y_i I_i)}{(\sum_k I_i)^2} = \frac{y_j - \bar{y}}{(\sum_k I_i)}, \]

\[ \frac{\delta \Sigma_{xx}}{\delta I_j} = \frac{[(x_j - \bar{x})^2 - 2(x_j - \bar{x}) \frac{\delta \bar{x}}{\delta I_j} I_j] \sum_k I_i - \sum_k [(x_i - \bar{x})^2 I_i]}{(\sum_k I_i)^2} \]

(3.41)

\[ \frac{\delta \Sigma_{yy}}{\delta I_j} = \frac{[(y_j - \bar{y})^2 - 2(y_j - \bar{y}) \frac{\delta \bar{y}}{\delta I_j} I_j] \sum_k I_i - \sum_k [(y_i - \bar{y})^2 I_i]}{(\sum_k I_i)^2} \]

(3.42)

\[ \frac{\delta \Sigma_{xy}}{\delta I_j} = \frac{[(x_j - \bar{x})(y_j - \bar{y}) - (y_j - \bar{y}) \frac{\delta \bar{x}}{\delta I_j} I_j] \sum_k I_i}{(\sum_k I_i)^2} \]

(3.43)
We can now compute the gradient of the second moment regularization within the minimization framework of the eht-imaging library:

$$\frac{\delta R_{\Sigma}}{\delta I_j} = 2(\Sigma_{xx} - \Sigma'_{xx}) \frac{\delta \Sigma_{xx}}{\delta I_j} + 2(\Sigma_{yy} - \Sigma'_{yy}) \frac{\delta \Sigma_{yy}}{\delta I_j} + 4(\Sigma_{xy} - \Sigma'_{xy}) \frac{\delta \Sigma_{xy}}{\delta I_j}. \quad (3.44)$$

Note that these equations correspond to regularization of the normalized second central moment of an image. In cases where the total flux density of an image is constrained or regularized, it would be advantageous to instead regularize the unnormalized second central moment, giving a substantially simplified and convex optimization problem.
Chapter 4

Persistent Non-Gaussian Structure in the Image of Sagittarius A* at 86 GHz


Abstract

Observations of the Galactic Center supermassive black hole Sagittarius A* (Sgr A*) with very long baseline interferometry (VLBI) are affected by interstellar scattering along our line of sight. At long radio observing wavelengths (\(\gtrsim 1\) cm), the scattering heavily dominates image morphology. At 3.5 mm (86 GHz), the intrinsic source structure is no longer sub-dominant to scattering, and thus the intrinsic emission from Sgr A* is resolvable with the Global Millimeter VLBI Array (GMVA). Long-baseline detections to the phased Atacama Large Millimeter/submillimeter Array (ALMA) in 2017 provided new constraints on the intrinsic and scattering properties of Sgr A*, but the stochastic nature of the scattering requires multiple observing epochs to reliably estimate its statistical properties. We present new observations with the GMVA+ALMA, taken in 2018, which confirm non-Gaussian structure in the scattered image seen in 2017. In particular, the ALMA–GBT baseline shows more flux density than expected for an anistropic Gaussian model, providing a tight constraint on the source size and an upper limit on the dissipation scale of interstellar turbulence. We find an intrinsic source extent along the minor axis of \(\sim 100\) \(\mu\)as both via extrapolation of longer wavelength scattering constraints and direct modeling of the 3.5 mm observations. Simultaneously fitting for the scattering parameters, we find an at-most modestly asymmetrical (major-to-minor axis ratio of \(1.5\pm0.2\)) intrinsic source morphology for Sgr A*. 
Chapter 4: Persistent Non-Gaussian Structure in the Image of Sagittarius A*

4.1 Introduction

The Galactic Center hosts the closest known supermassive black hole (SMBH), associated with the radio source Sagittarius A* [Sgr A*; Balick & Brown, 1974]. With a mass $M \sim 4.1 \times 10^6 M_\odot$ at a distance $D \sim 8.1$ kpc, Sgr A* subtends the largest angle on the sky among all known black holes [Ghez et al., 2008; Reid, 2009; Gillessen et al., 2009; Gravity Collaboration et al., 2018a]. Thus Sgr A* is one of the most promising targets to study black hole accretion and outflow via direct imaging [Goddi et al., 2017]. The spectral energy density of Sgr A* in radio rises with frequency, with a turnover in the sub-millimeter regime, where the accretion flow becomes optically thin [Falcke et al., 1998; Bower et al., 2015b, 2019]. However, the southern declination and interstellar scattering of Sgr A* add challenges to decades of radio observations with very long baseline interferometry [VLBI; Alberdi et al., 1993; Backer et al., 1993; Krichbaum et al., 1993; Marcaide et al., 1999; Bower et al., 2004; Shen et al., 2005; Lu et al., 2011a; Bower et al., 2014]. Thus, the intrinsic accretion and outflow structure of Sgr A* remains rather poorly understood.

Early observations at 1.3 mm with the prototype Event Horizon Telescope (EHT) indicate that the radio emission of Sgr A* originates from a region that is comparable to the size of the black hole’s “shadow” [∼50 µas; Doeleman et al., 2008; Fish et al., 2011; Johnson et al., 2015; Fish et al., 2016; Lu et al., 2018]. On these scales, the image morphology is dominated by strong gravitational lensing of the black hole rather than by details of the innermost accretion flow [as seen in M87; Event Horizon Telescope Collaboration et al., 2019a]. At longer wavelengths, images of Sgr A* are strongly scatter-broadened (blurred) by the intervening interstellar medium [ISM; e.g., Davies et al., 1976; van Langevelde et al., 1992; Frail et al., 1994; Bower et al., 2004; Shen et al., 2005; Bower et al., 2006; Psaltis et al., 2018; Johnson et al., 2018].

Radio waves passing through the ionized ISM propagate via multiple paths due to changes in the refractive index of the turbulent plasma from density inhomogeneities. The angles at which the waves scatter are proportional to the squared wavelength of the wave. The intrinsic angular size of Sgr A* at wavelengths of 0.1–1 centimeters is roughly proportional to the wavelength. As a result, the ratio of intrinsic source angular size to scatter-broadening is $\sim 0.3/\lambda_{\text{cm}}$ along the major axis and $\sim 0.6/\lambda_{\text{cm}}$ along the minor axis [where $\lambda_{\text{cm}}$ is the observing wavelength in centimeters; Johnson et al., 2018], making 3.5 mm the longest observing wavelength accessible on Earth at which Sgr A* intrinsic structure would not be sub-dominant to scattering. The ideal regime to probe and separate intrinsic source properties from scattering is thus at 3.5 mm: intrinsic structure starts to dominate over scattering effects, and the radio emission originates from the optically thick innermost accretion flow, also corresponding to the launching region of a possible outflow or jet [Narayan et al., 1995; Falcke & Markoff, 2000; Özel et al., 2000; Yuan et al., 2003].

Although intrinsic structure dominates at 1.3 mm, the scattering is nonetheless still substantial, and could potentially contaminate tests of general relativity with the EHT, introducing random distortions and substructure in the image. The specific effects on 1.3 mm VLBI images depend on the power spectrum $Q(q)$ of spatial irregularities that produce the scattering [where $q$ is a wavevector; Johnson et al., 2018; Zhu et al., 2019]. Because these underlying irregularities that cause refractive scattering at 3.5 mm also produce image variations at 1.3 mm, scattering studies at 3.5 mm are essential to guide imaging Sgr A* at 1.3 mm with the EHT. Furthermore, scattering-induced substructure, predicted by Narayan & Goodman [1989] and Goodman & Narayan [1989]
and first measured in Sgr A* by Gwinn et al. [2014] at 1.3 cm, is caused by modes in the scattering material on scales much larger than the diffractive scale of the scattering. This turbulence in the ISM induces stochastically varying compact substructure in images of Sgr A* that contaminates long-baseline source behavior with added “refractive noise” in the visibility domain, making the recovery of small-scale intrinsic source structure difficult [Johnson & Gwinn, 2015; Johnson & Narayan, 2016].

In this paper, we utilize the scattering model developed by Psaltis et al. [2018], using physical parameters from Johnson et al. [2018] that were estimated using archival observations of Sgr A*. The two-dimensional power spectrum of the phase fluctuations $Q(q)$ is modeled as an unbroken anisotropic power-law with a spectral index $\beta$ extending between a maximum scale (the outer scale $r_{\text{out}}$) and a minimum scale (the inner scale $r_{\text{in}}$): $Q(q) \propto |q|^{-\beta}$ ($\beta$ is also the exponent for the three-dimensional power spectrum of density fluctuations; e.g., Blandford & Narayan, 1985; Rickett, 1990). This power spectrum then yields a second-order phase structure function $D_{\phi}(r) = \langle (\phi(r + \mathbf{r}) - \phi(\mathbf{r}))^2 \rangle \propto |r|^\alpha$ in the inertial range $r_{\text{in}} \ll r \ll r_{\text{out}}$, where $\alpha \equiv \beta - 2$. While two scattering models may have identical scatter-broadening, they may still differ wildly in their refractive substructure. Combining information from both scatter-broadening from 1.3 mm to 30 cm and centimeter-wave substructure strongly constrains the scattering power spectrum and the asymptotic Gaussian morphology parameters of the scatter-broadening kernel. However, a degeneracy between the power-law index $\alpha$ and the inner scale of the turbulence in the ISM $r_{\text{in}}$ remains [Johnson et al., 2018]: various combinations of scattering and intrinsic source parameters can produce the same observed behavior in the scattered image, illustrated in Figure 4.1. Sensitive VLBI observations at 3.5 mm offer a prime opportunity to break degeneracies between the parameters by connecting to millimeter-wave scattering behavior.

For the past two decades since its first detection at 3.5 mm [Rogers et al., 1994], the scattered image of Sgr A* has been commonly modeled as an elliptical Gaussian source with a position angle of $\sim 80^\circ$ east of north [e.g., Shen et al., 2005; Bower et al., 2006; Lu et al., 2011a; Ortiz-León et al., 2016; Brinkerink et al., 2019]. Closure phases — the directed sums of visibility phases over closed triangles of baselines and robust to station-based errors [e.g, Cornwell, 1989; Rauch et al., 2016; Thompson et al., 2017; Blackburn et al., 2020] — measured by early 3.5 mm experiments were consistently zero [Rogers et al., 1994; Krichbaum et al., 1998; Doeleman et al., 2001; Shen et al., 2005; Bower et al., 2006; Lu et al., 2011a], indicating source symmetry on the probed spatial scales, while non-zero closure phases at lower frequencies are entirely attributable to interstellar scattering [Johnson et al., 2018]. A new set of higher-sensitivity experiments, including the Large Millimeter Telescope Alfonso Serrano (LMT) and the Robert C. Byrd Green Bank Telescope (GBT), detected the first non-zero 3.5 mm closure phases in Sgr A* on new closure triangles provided by the addition of the LMT, which could either be due to intrinsic structure or non-Gaussian structure in the scattering screen [Ortiz-León et al., 2016; Brinkerink et al., 2016]. These results motivate breaking the assumptions of an elliptical Gaussian source model and attempting to recover complex underlying source structure via imaging.

The recently added VLBI phasing capability to the Atacama Large Millimeter/submillimeter Array (ALMA) provided additional sensitivity and long north-south baselines to the Global Millimeter VLBI Array (GMVA) at 3.5 mm [Matthews et al., 2018]. While pre-ALMA experiments could not identify the detailed morphology or constrain the radio emission model, the sensitivity
Figure 4.1: Modeled effects of interstellar scattering on a simulated source whose intrinsic structure is that of a circular Gaussian with a full-width at half-maximum (FWHM) of 100 $\mu$as (shown as the solid circle) at $\lambda = 3.5$ mm with varying power-law index $\alpha$ and inner scale of turbulence $r_{\text{in}}$. The dashed ellipse shows the FWHM size of the measured Gaussian source on the sky for Sgr A* in previous 3.5 mm experiments [Ortiz-León et al., 2016; Brinkerink et al., 2019]. Each row has constant $\alpha$ and varying $r_{\text{in}}$ (10, $10^3$, and $10^5$ km): as $r_{\text{in}}$ increases, the scatter-broadening and the level of refractive substructure increase. Each column has constant $r_{\text{in}}$ and varying $\alpha$ (1.0, 1.45, and 1.9): as $\alpha$ increases, the scatter-broadening and the level of refractive substructure increase.
and coverage brought by joining ALMA to the GMVA for the first time in 2017 — including tripling the angular north-south resolution — has offered a major leap in imaging capabilities and model discrimination for Sgr A*.

In Issaoun et al. [2019b] (Chapter 2), we showed that measured visibility amplitudes on long baselines to ALMA exhibit clear non-Gaussian behavior, which was a function of baseline length, and ruled out a potential scattering model that would significantly contaminate future EHT images at 1.3 mm [Zhu et al., 2019]. Using interstellar scattering mitigation methods [Johnson, 2016] coupled with the enhanced coverage of GMVA+ALMA, we then reconstructed a first image of the unscattered structure of Sgr A* at 3.5 mm. The unscattered source had a major axis full width at half-maximum (FWHM) of $120^{\pm}34^{\mu}\text{as}$ ($12.0^{+3.4}_{-0.2}$ Schwarzschild radii ($R_{Sch}$); where $R_{Sch} = 2 GM/c^2$) and a circularly symmetric morphology (major-to-minor-axis ratio of $1.2^{+0.3}_{-0.2}$), which requires either that the accretion flow dominates the emission or that jet-dominant emission from Sgr A* is pointed within $20^{\circ}$ of the line of sight.

Issaoun et al. [2019b] used the scattering model that best matched observations [Johnson et al., 2018, hereafter model J18] in the imaging process to mitigate the scattering effects and recover the intrinsic structure. Based on a single observation, it is not clear whether the scattering parameter assumptions for Sgr A* are valid: long-baseline detections could either be attributed to intrinsic structure, scattering substructure, or a mix of both. The 2017 long-baseline detections were more consistent with the near-Kolmogorov power spectrum (power-law index $\alpha = 1.38$) from J18 rather than a flat spectrum [Goldreich & Sridhar, 2006, hereafter model GS06] governing the stochastic variations in the refractive noise. It is however possible that the low refractive noise observed can be attributed to a statistically unlikely low refractive noise realization from the GS06 flat-spectrum scattering model. Thus, it remains important to sample these scales at different instances in time. We therefore performed follow-up observations of Sgr A* with GMVA+ALMA in 2018 to gain further confidence in the scattering model and tighten constraints in the model parameters to describe the ISM along the line of sight to Sgr A*.

The organization of the paper is as follows. We summarize observations and data reduction in Section 4.2, present our final GMVA+ALMA visibility amplitudes on Sgr A* in Section 4.3, and discuss the constraints on scattering and intrinsic source parameters enabled by these latest results in Section 4.4. A summary is given in Section 4.5.

### 4.2 Observations and data reduction

We observed Sgr A* ($\alpha_{J2000} = 17^{h}45^{m}40^{s}.0361$, $\delta_{J2000} = -29^{\circ}00^{\prime}28^{\prime\prime}.168^{1}$) with the GMVA (project code MJ001), composed of the eight Very Long Baseline Array (VLBA) antennas equipped with 86 GHz receivers, the GBT, and phasing 35 single ALMA antennas [Matthews et al., 2018]. The observations were conducted on 14 and 17 April 2018 as part of the Cycle 5 VLBI session with ALMA (project code 2017.1.00795.V). GMVA stations recorded a total bandwidth of 256 MHz per polarization divided into eight 32 MHz-wide intermediate frequencies (IFs), while ALMA recorded overlapping 62.6 MHz IFs separated by 58.59375 MHz that fully covered the GMVA band. The two 6 hr tracks included three calibrator sources: 3C279, NRAO 530, and J1924–2914. The GBT

---

1Coordinates from the NRAO VLBI observing schedules C181B and C181F are provided by the NRAO SCHED program: http://www.aoc.nrao.edu/~cwalker/sched/Source_Catalog.html

87
Figure 4.2: Top: The Sgr A* \((u,v)\) coverage, showing non-detections in gray, and detections for 14 April (blue) and 17 April 2018 (red). Each symbol denotes a scan-averaged (over \(\sim 9\) minutes) measurement. Bottom: The signal-to-noise ratio (S/N) for scan-averaged visibilities on Sgr A* as a function of projected baseline length, showing detections for 14 April 2018 (blue) and 17 April 2018 (red). The gray dashed line in both panels delimits baseline lengths equivalent to a resolution of 200 \(\mu\)as. All detections beyond \(~1\)G\(\lambda\) are on baselines to ALMA. 17 April has fewer long-baseline detections due to the absence of the GBT to anchor the fringe calibration of the array, caused by a recording disk failure at the correlation stage.
4.3 Results

participated for 3 hours in each track, but the station recording was lost on 17 April due to a recording disk failure.

The data were correlated with the VLBI correlator at the Max Planck Institute for Radio Astronomy in Bonn using the DiFX software [Deller et al., 2011]. To accommodate the noncongruent IF configuration between ALMA and the other stations, data were correlated over distinct sub-IFs and synthesized back into contiguous GMVA IFs using DiFX tool difx2difx. Mixed linear-circular polarization correlation products between ALMA and the GMVA were transformed to pure circular polarization via PolConvert [Martí-Vidal et al., 2016b], utilizing a full calibration of the ALMA interferometric products performed by the ALMA quality assurance (QA2) team [Goddi et al., 2019]. Data were then fringe fitted and reduced using the enhanced Haystack Observatory Postprocessing System\(^2\) pipeline (EHT-HOPS) presented in Blackburn et al. [2019], with additional validation and cross-checks from the NRAO Astronomical Image Processing System [AIPS; Greisen, 2003]. The EHT-HOPS pipeline introduces a number of key improvements over the original HOPS software, including global fringe fitting and improved phase calibration. Our 86 GHz implementation of the EHT-HOPS reduction follows the same procedure as in Issaoun et al. [2019b]. We performed a priori amplitude calibration and opacity correction with the task APCAL within AIPS, using observatory-provided telescope gain information and system temperatures measured during the observations. To form Stokes \(I\), corrections for field angle rotation and polarimetric gain ratios between left and right polarizations were derived and applied using the EHT Analysis Toolkit\(^3\) (eat library) polarimetric calibration framework for the EHT-HOPS pipeline [Blackburn et al., 2019; Steel et al., 2019].

Figure 4.2 shows the detections and non-detections per scan of \(\sim 9\) minutes for Sgr \(A^*\) (top panel) and corresponding signal-to-noise ratio (S/N) of scan-averaged visibilities for Sgr \(A^*\) detections for both observed epochs. All detections beyond \(\sim 1 \text{G} \lambda\) are on baselines to ALMA. Fringe solutions (delays and delay-rates) are determined from detections with S/N > 7 over the scan, and visibilities on weaker baselines can be measured once station delays and delay-rates are known. After a priori calibration, we proceed to further calibrate antenna gains from calibrator imaging and recover improved measurements of the visibility amplitudes for Sgr \(A^*\).

4.3 Results

Sensitivity estimates from a priori amplitude calibration of individual telescopes commonly overestimate station performance, not taking into account effects such as pointing and focus errors, receiver misalignments, operational difficulties, or unstable weather conditions. The visibility amplitudes obtained on Sgr \(A^*\) from a priori calibration alone do not fully capture true source behavior, and further steps must be taken to disentangle station-based residual amplitude errors from source signal before any more detailed analysis can take place. The amplitude calibration for Sgr \(A^*\) is done in three stages:

1. a priori amplitude calibration using site metadata (see Section 4.2);
2. inner 1 \text{G} \lambda self-calibration to a Gaussian source (see Section 4.3.2);

\(2\)https://www.haystack.mit.edu/tech/vlbi/hops.html
\(3\)https://github.com/sao-eht/eat
Chapter 4: Persistent Non-Gaussian Structure in the Image of Sagittarius A* 

3. residual self-calibration in THEMIS modeling of the intrinsic and scattering parameters (see Section 4.4.2).

We present our calibrator imaging in Section 4.3.1, our calibration methods for Sgr A* data and final gain constraints in Section 4.3.2, and final visibility amplitudes in Section 4.3.3.

4.3.1 Imaging the Calibrator J1924–2914

To improve the amplitude calibration of the array, we utilize J1924–2914, one of the two calibrators observed alternating with Sgr A*, and with sufficient coverage for imaging (see Figure 4.3). J1924–2914 appears point-like to ALMA as a connected-element interferometer, and its source-integrated flux density is measured by interferometric-ALMA observations simultaneous to our

Table 4.1: Source-integrated flux density of observed sources from interferometric-ALMA.

<table>
<thead>
<tr>
<th>Source</th>
<th>14 April $S_\nu$ (Jy)</th>
<th>17 April $S_\nu$ (Jy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sgr A*</td>
<td>2.2 ± 0.2</td>
<td>2.3 ± 0.2</td>
</tr>
<tr>
<td>NRAO 530</td>
<td>3.2 ± 0.3</td>
<td>3.2 ± 0.3</td>
</tr>
<tr>
<td>J1924–2914</td>
<td>4.6 ± 0.5</td>
<td>4.5 ± 0.5</td>
</tr>
<tr>
<td>3C 279</td>
<td>14 ± 1</td>
<td>15 ± 1</td>
</tr>
</tbody>
</table>

NOTE— These values are provided with interferometric-ALMA data as part of the ALMA Quality Assurance 2 (QA2) process, and assume a 10% uncertainty in the ALMA amplitude calibration [Goddi et al., 2019].
4.3 Results

Figure 4.4: Closure-only images and elliptical Gaussian component model fits of J1924–2914 using the eht-imaging library. The contour levels are 10–90% of the peak in steps of 10%. Black markers denote the central location of each model component in the model fits, produced using the eht-modeling module of the library. The images and model fits are restored with a circular Gaussian beam with a FWHM = 1/\(u_{\text{max}}\) = 83 \(\mu\)as for 14 April and 1/\(u_{\text{max}}\) = 76 \(\mu\)as for 17 April. The restoring beam is shown in the lower right corner of each plot. All images and model-fits have reduced \(\chi^2\) < 1.7 on closure products with no systematic noise budget added.

VLBI tracks (see Table 4.1). Its compactness, stable structure even on VLBI scales within a single epoch, and known flux density make it an ideal imaging target to obtain an estimate of station-based residual gain corrections during our observations via self-calibration.

While having \((u,v)\) coverage comparable to J1924–2914, the other imageable calibrator, NRAO530, is a north-south extended source [Bower et al., 1997; Bower & Backer, 1998; Feng et al., 2006; Chen et al., 2010; Lu et al., 2011c; Brinkerink et al., 2019; Issaoun et al., 2019b]. In an array configuration where the only north-south baselines resolving the jet are long baselines to ALMA, imaging of the source with the full array proved very difficult, resulting in unphysical images severely impacted by the lack of intermediate baselines between inter-VLBA and ALMA baselines. Imaging without ALMA resulted in various source structures (and varied gain corrections) that provided statistically acceptable fits to closure quantities, leading to low confidence in any single image structure. We thus omit NRAO530 from our gain analysis. We also observed 3C279 as a fringe-finder source with the full array for a few minutes but omit it from further analysis because its \((u,v)\) coverage is insufficient for imaging.

Following the same method as Issaoun et al. [2019b], we made use of the large number of closure phases and log closure amplitudes — constructed via the quotient of two visibility amplitude products involving four stations and robust to station-based errors [Twiss et al., 1960; Readhead et al., 1980] — and the total flux density constrained by the interferometric ALMA measurement in each observation to image J1924–2914 with the eht-imaging library [Chael et al., 2016, 2018b]. Closure-only imaging is robust to station-based instrumental errors, allowing us to reconstruct source morphology and derive residual telescope gain corrections directly from self-calibration to the obtained brightness distribution. We confirm general image morphology and gain trends via model fitting to the observed closure products with simple elliptical Gaussian components, and we test goodness-of-fit by calculating a reduced \(\chi^2\) of the model prediction against measurements of closure phases and log closure amplitudes.
Table 4.2: Station median multiplicative gain corrections to the visibility amplitudes for J1924–2914 compared to Sgr A*.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-Cal</td>
<td>Model Fit</td>
<td>Image</td>
<td>Self-Cal</td>
<td>Model Fit</td>
<td>Image</td>
</tr>
<tr>
<td>ALMA</td>
<td>—</td>
<td>1.8±0.4</td>
<td>1.3±0.2</td>
<td>—</td>
<td>1.1±1.2</td>
<td>1.0±1.1</td>
</tr>
<tr>
<td>BR</td>
<td>2.4±1.1</td>
<td>1.9±2.7</td>
<td>2.0±2.9</td>
<td>1.6±0.5</td>
<td>1.6±2.8</td>
<td>1.6±2.7</td>
</tr>
<tr>
<td>FD</td>
<td>2.2±1.0</td>
<td>2.6±2.3</td>
<td>2.8±2.3</td>
<td>1.9±0.2</td>
<td>3.4±3.4</td>
<td>3.3±3.1</td>
</tr>
<tr>
<td>GBT</td>
<td>1.4±1.6</td>
<td>3.1±7.6</td>
<td>3.5±5.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>KP</td>
<td>2.0±1.5</td>
<td>2.0±4.2</td>
<td>2.0±4.2</td>
<td>2.3±1.3</td>
<td>2.9±0.6</td>
<td>2.9±0.6</td>
</tr>
<tr>
<td>LA</td>
<td>1.3±0.5</td>
<td>1.2±0.3</td>
<td>1.2±0.2</td>
<td>1.5±0.5</td>
<td>1.3±0.4</td>
<td>1.3±0.4</td>
</tr>
<tr>
<td>MK</td>
<td>—</td>
<td>3.7±3.7</td>
<td>5.7±2.7</td>
<td>—</td>
<td>5.3±1.9</td>
<td>3.7±1.1</td>
</tr>
<tr>
<td>NL</td>
<td>4.7±0.0</td>
<td>4.2±1.3</td>
<td>4.3±1.3</td>
<td>2.3±5.5</td>
<td>2.6±2.3</td>
<td>2.6±2.3</td>
</tr>
<tr>
<td>OV</td>
<td>1.8±0.4</td>
<td>1.4±0.4</td>
<td>1.4±0.4</td>
<td>1.9±0.3</td>
<td>1.7±0.3</td>
<td>1.8±0.4</td>
</tr>
<tr>
<td>PT</td>
<td>2.0±0.7</td>
<td>3.6±1.8</td>
<td>3.5±1.8</td>
<td>2.0±1.2</td>
<td>4.0±1.4</td>
<td>3.9±1.5</td>
</tr>
</tbody>
</table>

NOTE – Median multiplicative gain corrections (and 95% interval of variation over time) to the visibility amplitudes for common stations from the two calibration methods: 1) self-calibration of Sgr A* amplitudes below 1Gά to the Gaussian source estimated from O16: B19, 2) self-calibration of J1924–2914 observations to the closure-only images, and 3) self-calibration of J1924–2914 observations to the closure-only model fits.

We present J1924–2914 images and model fits\(^4\) for both epochs in Figure 4.4. The observations have a uniform-weighted beam = (128×83) μas and PA = 38° for 14 April, and a uniform-weighted beam = (129×76) μas and PA = 45° for 17 April. The resulting images and model fits have reduced χ² < 1.7 on closure products with no error budget inflation added to account for possible systematics. A systematic error budget of 1% (1% of amplitudes added in quadrature to the thermal noise on complex visibilities) is required to drive reduced χ² to unity. The location of the components agree well between the images and model fits of each respective observation. The 14 April morphology is best described with four elliptical Gaussian components based on the lowest reduced χ² on the closure data products and realistic station gain reconstructions, while the 17 April morphology, missing GBT baselines, is best constrained with three elliptical Gaussian components. On both days, observations indicate a consistent north-west source elongation. The north-west elongation of the J1924–2914 jet is consistent with our first image of the source at 3.5 mm [Issaoun et al., 2019b] and follows the mm-jet morphology from 7 mm [Shen et al., 2002] and 1.3 mm observations [Lu et al., 2012].

4.3.2 Calibrating Sagittarius A* Visibility Amplitudes

Because of its scatter-broadening, we have fewer detections and lower S/N for Sgr A* than for the calibrators. Our data sets thus do not have sufficiently robust closure quantities to drive the

\(^4\)Throughout this paper we use perceptually uniform colormaps from the ehtplot library, https://github.com/rndsrc/ehtplot.
4.3 Results

recovery of non-trivial structure in closure-only imaging. Following the methodology employed in Issaoun et al. [2019b], we use two methods for amplitude calibration:

1. we obtain station gain trends from self-calibration to closure-only model fits and images of a calibrator;
2. we obtain station gain trends directly from Sgr A* by self-calibrating all visibility amplitudes within 1 Gλ using an elliptical Gaussian visibility function obtained from previous 3.5 mm experiments [Ortiz-León et al., 2016; Brinkerink et al., 2019, hereafter O16, B19 respectively].

For the second method, we assume that the behavior of the visibility amplitude function for Sgr A* on baselines within 1 Gλ is dominated by the image second moment [e.g., Hu, 1962; Issaoun et al., 2019a, see Chapter 3], which follows that of the visibility function of a Gaussian source with a FWHM size of 215 by 140 µas and a position angle of 80° (east of north). The choice of Gaussian widths is motivated by measurements from previous 3.5 mm experiments that included the sensitive LMT improving the recovery of the minor axis size [O16; J18; B19]. We therefore self-calibrated our Sgr A* amplitudes within 1 Gλ to the expected Gaussian morphology and obtain station-based amplitude gain corrections that are subsequently applied to correct visibility amplitudes on all baselines on a time-varying point-by-point basis. Since ALMA does not have any baselines within the 1 Gλ cutoff, it cannot be calibrated via this method.

In Table 4.2, we present the median multiplicative station gain corrections obtained via imaging and model fitting of J1924−2914 and short-baseline (within 1 Gλ) self-calibration of Sgr A* to an expected Gaussian source size. The J1924−2914 imaging/modeling and the Sgr A* self-calibration methods gave comparable gain solutions for most stations, validating the Gaussian source assumption for short-baseline measurements of Sgr A*.

According to the VLBA logs, North Liberty (NL) and Mauna Kea (MK) had poor weather on 14 April, which is consistent with the high gain corrections found for these two stations. For 17 April, NL gain corrections are not well constrained as its shortest baseline, to the GBT, is missing. For subsequent analysis, we flag NL for all data sets. MK is not present in the Sgr A* data set, as the source is too scatter-broadened to be detected on long baselines to MK. It is worth noting that we tend to recover higher gain corrections from the J1924−2914 Gaussian component model fitting than from the direct imaging for the stations with only long baselines (ALMA, MK), as the model fits do not capture smaller structural variations. Given the good consistency between imaging and modeling gain corrections for all other stations, we adopt imaging gain corrections for long-baseline stations. Note that we apply all derived gain corrections as a function of time to the data, not just the median scaling presented in Table 4.2. Because ALMA’s multiplicative gain corrections are near unity for J1924−2914, both for imaging and modeling, applying them would not significantly change the flux density on ALMA baselines. We thus choose not to apply ALMA gain corrections as not to introduce scatter from the gain solutions to the visibility amplitudes.

The stations with significant discrepancies between gain solutions derived from Sgr A* and J1924−2914 are GBT and Pie Town (PT) for 14 April, and Fort Davis (FD) and PT for 17 April. Both FD and PT have other VLBA stations very close to them, which allow them to be well-constrained by the Gaussian source assumption for Sgr A* whereas some extended diffuse features may be missing in the imaging/modeling of J1924−2914 that lead to this discrepancy.

93
Figure 4.5: GBT multiplicative 10-second interval gain correction trends for J1924–2914 (from the closure-only image) and Sgr A* (from self-calibration to a Gaussian source size) derived from the 14 April observations. The elevations of both sources are descending during the GBT observing track, and remain below 25° for all scans. Vertical dotted lines denote the times at which the GBT performed pointing scans on calibrator sources. The intra-scan scatter is likely due to pointing errors. High gain corrections after 13 UT and large scatter are likely due to faulty re-pointing or non-optimal surface adjustment of the telescope, solely affecting half of the J1924–2914 scans.
4.3 Results

Figure 4.6: Noise-debiased correlated flux density of Sgr A* as a function of projected baseline length for data after self-calibrating to the Gaussian source size from O16 and B19 using only baselines shorter than 1 Gλ. The errorbars indicate the thermal error in individual scans. The amplitude uncertainties from imperfect calibration are not shown but are of the order of ~ 10%, corresponding to roughly the size of a symbol. Three epochs are depicted: 3 April 2017 observations presented in Issaoun et al. [2019b] are shown in black; 14 April 2018 observations are shown in blue; and 17 April 2018 observations are shown in red (no GBT). Baselines to NL are flagged for all data sets due to erratic amplitude gain corrections from bad weather. Dashed and dotted curves show the expected visibilities along the major and minor axes, respectively, for an intrinsic elliptical Gaussian source with a FWHM size of $140 \times 105 \mu$as scattered with two scattering models: in magenta the intrinsic Gaussian source is scattered with the estimated J18 scattering parameters $\alpha = 1.38$ and $r_{\text{in}} = 800 \text{ km}$; while in black the same Gaussian source is scattered with a Gaussian scattering kernel ($169 \times 86 \mu$as, position angle of 81.9° east of north; J18), with $r_{\text{in}} \to \infty$. The orange and green curves show the expected RMS renormalized refractive noise along the major and minor axes for the GS06 and J18 models respectively. All detections beyond ~1Gλ are baselines to ALMA oriented close to the scattering minor axis. Labeled filled triangles, colored by data set, indicate 4σ upper limits on four sensitive baselines at other orientations, where corresponding detections were found for our calibrators. Detections on the ALMA–GBT baseline (filled diamonds), oriented along the scattering minor axis, sit above the expected flux density for a Gaussian source and thus clearly indicate non-Gaussian source morphology. Detections on baselines beyond 2Gλ exhibit expected properties of refractive noise. For both years, refractive noise detections match the average level predicted by J18 and sit below that of GS06.
Since the ALMA–GBT baseline is crucial to understanding deviations from the Gaussian source assumption for Sgr A*, we take a closer look at the derived gain corrections for the GBT using both methods. In Figure 4.5, we show the 14 April GBT gain trends for the Sgr A* Gaussian source method and the J1924–2914 image as a function of time. The discrepancy between the two sources is due to a systematic offset for half of the GBT scans on J1924–2914 at the end of the GBT track. This offset is possibly due to a faulty pointing solution or non-optimal surface adjustment for the telescope after 13 UT affecting half of the J1924–2914 scans. For the times where both sources are observed intermittently, there is a good agreement between the derived gain corrections, with mean GBT amplitude gain corrections of $1.25 \pm 0.08$ and $1.2 \pm 0.1$ for J1924–2914 and Sgr A* respectively. We thus choose to proceed with the Gaussian source-derived gain corrections for the calibration of Sgr A* GBT baselines.

### 4.3.3 Final Sgr A* Visibility Amplitudes

In Figure 4.6, we show the scan-averaged noise-debiased visibility amplitudes for Sgr A* after Gaussian source self-calibration of the inner 1G$\lambda$ for all GMVA+ALMA observations to date (3 April 2017, 14 April 2018, 17 April 2018). For noise-debiasing, thermal noise contributions to visibility amplitude are removed according to the prescription in Johnson et al. [2015] [see also Thompson et al., 2017]. The ALMA–GBT baseline, sitting along the scattering minor axis, gives significantly higher flux density than that expected for the minor axis of an intrinsic Gaussian source with a FWHM size of $140 \times 105 \mu$as scattered by a purely Gaussian scattering screen ($r_{\text{in}} \to \infty$; black curves) but matches the expected flux density from an intrinsic Gaussian source of the same angular size scattered with the estimated parameters from the J18 scattering model ($r_{\text{in}} = 800$ km; magenta curves). In 2018, VLBA detections to ALMA, oriented near the scattering minor axis, show clear deviations from Gaussian behavior for the scattered image of Sgr A* on the sky and exhibit properties expected from refractive noise. We also derive 4$\sigma$ upper limits on long baselines with other orientations, based on S/N for detections of the calibrators (filled triangles in Figure 4.6). Thus our new observations exhibit deviations from Gaussian morphology similar to those of our 2017 observations presented in Issaoun et al. [2019b]. For both 2017 and 2018, the ALMA–GBT baseline exhibits a flux density excess and long-baseline VLBA detections to ALMA are consistent with the average refractive noise predicted by J18, sitting below that of GS06. These two scattering realizations one year apart allow us to confidently exclude GS06 as a viable model for the interstellar scattering along our line of sight toward Sgr A*.

The coverage of our 2018 observations, with only 3 hours including GBT in one epoch, is insufficient to image Sgr A* with the techniques used for the 2017 data set [Issaoun et al., 2019a,b]. The large amplitude uncertainty on short VLBA baselines and the lack of non-zero closure phases prevent high-fidelity imaging. A wide variety of image morphologies can be obtained with similar goodness-of-fit to the data, therefore any images we obtain would be driven by strong prior assumptions in the imaging process. However, even without imaging, a direct analysis of the visibility amplitudes can provide strong constraints on the intrinsic size of Sgr A* and the inner scale of the interstellar scattering.
### Table 4.3: Results from THEMIS model fitting of the intrinsic source and scattering parameters simultaneously.

<table>
<thead>
<tr>
<th>Variable</th>
<th>3 Apr 2017</th>
<th>14 Apr 2018</th>
<th>17 Apr 2018</th>
<th>J18</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{\text{min}}$ (mas)</td>
<td>99$^{+7}_{-6}$</td>
<td>96$^{+7}_{-6}$</td>
<td>110$^{+40}_{-52}$</td>
<td>—</td>
</tr>
<tr>
<td>$\theta_{\text{maj}}$ (mas)</td>
<td>146$^{+11}_{-12}$</td>
<td>140$^{+14}_{-13}$</td>
<td>129$^{+67}_{-29}$</td>
<td>—</td>
</tr>
<tr>
<td>$\phi$ (deg.)</td>
<td>70$^{+6}_{-6}$</td>
<td>66$^{+16}_{-9}$</td>
<td>72$^{+95}_{-63}$</td>
<td>—</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.8$^{+0.8}_{-0.7}$</td>
<td>1.1$^{+0.8}_{-1.0}$</td>
<td>1.0$^{+0.9}_{-0.9}$</td>
<td>1.38$^{+0.08}_{-0.04}$</td>
</tr>
<tr>
<td>$\log_{10}(r_{\text{in}}/\text{1 km})$</td>
<td>3.2$^{+0.5}_{-0.6}$</td>
<td>5$^{+2}_{-2}$</td>
<td>4$^{+3}_{-2}$</td>
<td>2.9$^{+0.1}_{-0.1}$</td>
</tr>
</tbody>
</table>

**NOTE**—Median values and 95% confidence ranges are shown. The posteriors have been filtered to remove numerical pathologies at the edges of the sampled scattering parameter ranges.

### 4.4 Discussion

Connecting scattering properties observed at centimeter wavelengths to those at millimeter wavelengths is not trivial, and depends on the relationship between the diffractive scale of the scattering $r_{\text{diff}} \sim \lambda/\theta_{\text{scatt}}$ and the dissipation scale of turbulence in the scattering material [Psaltis et al., 2018; Johnson et al., 2018]. Despite the precise measurement of the angular broadening size of Sgr A* at centimeter wavelengths, the scattering properties cannot be well constrained without a good understanding of the expected transition to non-$\lambda^2$ and non-Gaussian scattering at millimeter wavelengths, where the dissipation scale is comparable to the diffractive scale. The observations presented in this paper offer a prime opportunity to break the degeneracies between the scattering parameters.

We assume a single thin scattering screen incorporating observed behavior from long and short wavelength observations. The parameterization of the uncertainties in the scattering properties is motivated with physical models of the ISM material. These models are typically an anisotropic power-law with an index $\alpha$ for the power spectrum governing phase variations, extending between a maximum (or outer) scale $r_{\text{out}}$ and a minimum (or inner) scale $r_{\text{in}}$. We utilize the scattering model developed by Psaltis et al. [2018], assuming the “dipole” model for the magnetic field wander. The parameters that can be varied are the asymptotic Gaussian source parameters of the scatter-broadening kernel ($\theta_{\text{maj},0}, \theta_{\text{min},0}, \text{and } \phi_{\text{PA}}$), the power-law index $\alpha$, and the inner scale of the turbulence $r_{\text{in}}$ that cause diffractive kernel deviations from Gaussian and $\lambda^2$ behavior.

In this Section, we fix the asymptotic Gaussian source parameters of the scatter-broadening kernel and the distances between the screen, the observer and Sgr A* to the well-constrained values derived from long-wavelength observations for the J18 model. In Section 4.4.1, we provide qualitative constraints on $\alpha$, $r_{\text{in}}$ and source extent using only long-wavelength constraints and the ALMA–GBT measurements at 3.5 mm, and in Section 4.4.2 we present a full modeling of the scattering and intrinsic source parameters. In Section 4.4.3, we discriminate between scattering models based on the expected and measured power on long baselines in our 3.5 mm data sets.
Figure 4.7: Constraints on $\alpha$ and $r_{\text{in}}$ as a function of intrinsic source extent that would result in the measured mean amplitude ($58 \pm 6 \text{ mJy}$) at $\lambda = 3.5 \text{ mm}$ on a projected baseline length of $1.6 \text{ G}\lambda$ along the scattering minor axis. The blue bands show the allowed ranges of $\alpha$ and $r_{\text{in}}$ for a given intrinsic source size along the scattering minor axis with the measured amplitude at the chosen baseline length. Source sizes in grayscale lie beyond the allowed $\pm 1\sigma$ intrinsic source FWHM estimates from historical $3.5 \text{ mm}$ measurements (106–121 $\mu$as; O16; J18; B19; Issaoun et al. 2019b). The red shaded region shows the range of $\alpha$ and $r_{\text{in}}$ constrained via $1.3 \text{ cm}$ and $7 \text{ mm}$ observations in Johnson et al. [2018].

Figure 4.8: Same as Figure 4.7, but with the added THEMIS posterior distributions from simultaneous intrinsic source and scattering parameter fitting overlaid in green. The $1\sigma$, $2\sigma$, and $3\sigma$ regions are shown from dark green to light green. The solid, dashed, and dotted lines are the $1\sigma$, $2\sigma$, and $3\sigma$ contours respectively.
4.4 Discussion

### 4.4.1 Long-Wavelength Constraints on Inner Scale and Intrinsic Size

In Figure 4.7, we present the current constraints on the scattering parameters \( \alpha \) and \( r_{\text{in}} \) for various intrinsic source extents derived from 3.5 mm measurements. Our measurements on the ALMA–GBT baseline, oriented along the scattering minor axis and resolving the source, indicate clear non-Gaussian source morphology that is persistent over two years of observations. Our historical size measurements at 3.5 mm constrain the intrinsic source FWHM along the scattering minor axis to 106–121 \( \mu \)as within 1 \( \sigma \) uncertainties, assuming the J18 scattering model [Johnson et al., 2018; Issaoun et al., 2019b]. We measured the combined 2017–2018 mean amplitude in the middle of the ALMA–GBT baseline track to be 58 \( \pm \) 6 mJy at a projected baseline length of 1.6 \( \lambda \), assuming a 10% uncertainty on the overall amplitude calibration.

As shown in Figure 4.6, low values for \( r_{\text{in}} \) lead to a shallower fall-off of the visibility amplitudes as a function of baseline length, while \( r_{\text{in}} \to \infty \) approaches perfect Gaussian behavior. Deviations from a Gaussian behavior can also be achieved with low values of \( \alpha \). Therefore, for a measured flux density on a given baseline probing the ensemble-average (purely scatter-broadened) image (with a given intrinsic source size, \( \alpha \), and \( r_{\text{in}} \)) the same flux density can be achieved with lower \( \alpha \) and \( r_{\text{in}} \) values paired with a larger intrinsic source. The blue shaded regions in Figure 4.7 thus show the ranges of parameters giving the measured flux density on ALMA–GBT for a given intrinsic source extent along the scattering minor axis at 3.5 mm within our 1 \( \sigma \) measured range assuming the J18 model. The gray shaded regions give examples of source extent that are beyond our 1 \( \sigma \) measured range.

The red shaded regions in Figure 4.7 show composite constraints from longer-wavelength radio observations (see Figure 9 of Johnson et al. 2018). The light red shaded region corresponds to the composite (95% confidence) ranges of \( \alpha \) and \( r_{\text{in}} \) able to reproduce the refractive noise measurements at both \( \lambda = 3.6 \text{ cm} \) and \( \lambda = 1.3 \text{ cm} \) but unable to reproduce the non-Gaussian source morphology observed at 7 mm, which requires an inner scale \( r_{\text{in}} < 2000 \text{ km} \). The dark red shaded region corresponds to the ranges of \( \alpha \) and \( r_{\text{in}} \) able to both reproduce the cm-wave refractive noise measurements and the non-Gaussian shape at \( \lambda = 7 \text{ mm} \). The lower limit of the red region corresponds to a lower limit of \( r_{\text{in}} \geq 520 \text{ km} \) constrained by the Gaussian source morphology at 1.3 cm. Further details on these longer-wavelengths constraints are presented in Johnson et al. [2018]. From the composite longer-wavelength model constraints and 3.5 mm size constraints in Figure 4.7, we conclude that the intrinsic source extent (FWHM) of Sgr A* along the scattering minor axis must be 100 – 105 \( \mu \)as. Future analysis of more recent longer-wavelength observations of Sgr A* can also further constrain the parameter space of the scattering model (I. Cho et al. in prep.).

### 4.4.2 Joint Modeling of Scattering and Source Parameters

Using only the 3.5 mm observations, we additionally simultaneously model the intrinsic source and interstellar scattering to obtain constraints on source and scattering parameters within the modeling and analysis framework THEMIS [Broderick et al., 2020]. THEMIS provides a number of methods for handling data, defining models, addressing data systematic uncertainties, and sampling the resulting likelihoods. Because the model has closure phases that are identically zero (as is statistically consistent with our data), we fit only the 3.5 mm final visibility amplitudes
obtained in Section 4.3.3, excluding those data points with an S/N less than 2 to avoid non-Gaussian errors [Thompson et al., 2017].

The primary output of THEMIS-based analyses are posteriors on model parameters implied by the input data. In this instance, to ensure global convergence, we use the parallelly-tempered Markov Chain Monte Carlo (MCMC) sampler: we employed the deterministic even-odd swap tempering scheme of Syed et al. [2019] with the automated factor slice sampler of Tibbits et al. [2014], which is very efficient for small numbers of model parameters.

Sgr A* is modeled as an elliptical Gaussian source convolved with a parameterized version of the anisotropic diffractive kernel described in Psaltis et al. [2018] and Johnson et al. [2018]. The elliptical Gaussian source is parameterized in terms of a total flux density, averaged size, axial ratio (major-to-minor axis ratio), and position angle ($\phi$) as described in Broderick et al. [2020], each with uninformative uniform priors. From these we construct the intrinsic source Gaussian major/minor axes ($\theta_{\text{maj}}, \theta_{\text{min}}$). Only the inner scale $r_{\text{in}}$ and power-law index $\alpha$ are permitted to vary, with the remaining scattering parameters already well constrained by prior data. For $r_{\text{in}}$ a log-uniform prior is assumed, ranging from 1 km to $10^7$ km; for $\alpha$ a uniform prior is assumed on (0, 2).

To accommodate the refractive scattering component a constant 7 mJy noise floor is added in quadrature to the input data uncertainties; this is both conceptually and computationally much simpler than fully modeling the complex phase screen. Station residual multiplicative gain corrections are reconstructed and marginalized over via the Laplace approximation during the construction of log-likelihoods, after imposing a Gaussian prior centered on unity with a standard deviation of 20% [see Section 6.8 of Broderick et al., 2020]. Fitting gain amplitudes has an added practical benefit: when fitting mock data sets, produced with similar noise properties and baseline coverage to our observations of Sgr A*, we found that fitting the gain amplitudes effectively mitigated biases from refractive scattering. The derived residual multiplicative gain corrections are shown in Figure 4.9. Each data set was independently analyzed to avoid potential complications associated with source variation. Excellent fits were obtained in all cases. No qualitative differences in fit quality were found when analyzing both 2018 data sets together.

In Figure 4.8, we present the posteriors on the scattering parameters $r_{\text{in}}$ and $\alpha$ added to the multi-wavelength analysis from Figure 4.7. The median values and 95% confidence ranges for all fitted parameters are presented in Table 4.3 for each data set. The results for 3 April 2017 and 14 April 2018 show great consistency within our confidence ranges, while the results for 17 April 2018 are not well-constrained due to the absence of the GBT. Although $\alpha$ is not well-constrained at 3.5 mm, the 2017 data set in particular constrains $r_{\text{in}} < 3300$ km at the 2σ level. The 2017 data set contains the longest observation with the ALMA–GBT baseline, on which the amplitude fall-off indicates deviations from Gaussian morphology and thus results in a finite inner scale, as illustrated in Figure 4.6. This result is completely independent from the longer-wavelength constraints (red shaded regions in Figure 4.8), yet provides complete overlap in the parameter space for $\alpha$ and $r_{\text{in}}$, as well as a consistent intrinsic size estimate to that derived from longer-wavelength constraints in Section 4.4.1. While the 2018 data sets alone do not constrain in the $\alpha - r_{\text{in}}$ parameter space, the clear ALMA–GBT detections on 14 April 2018 build confidence in the non-Gaussian behavior on that baseline across multiple years. The persistently low-flux-density long-baseline VLBA–ALMA detections allow us to rule out the GS06 scattering model as a viable
4.4 Discussion

**FIGURE 4.9:** Residual multiplicative gain corrections derived by THEMIS in the Sgr A* model fitting for all three observing days. Top: Residual variations in the amplitudes derived from the data sets where only a priori amplitude corrections are applied. Bottom: Residual variations in the amplitudes derived from the data sets already corrected based on self-calibration (see Section 4.3.2). For 3 April 2017, inner 1Gλ self-cal gain corrections were not applied to the GBT due to the scatter introduced by its baseline to NL, whose measurements dominate within that baseline cut [Issaoun et al., 2019b]. In all panels, individual station gain corrections are offset vertically by unity for clarity, with the dashed horizontal lines indicating a unit gain for each station. The gain correction scale is linear. Each point corresponds to an individual scan: colored points are independently reconstructed in the procedure, grey points are not well constrained due to missing information and are thus heavily biased by the prior.
model for the interstellar scattering toward Sgr A*, see Section 4.4.3.

We select the 3 April 2017 data set as our best data set, due to its more complete \((u,v)\) coverage and its longer GBT observing track (double that of the 2018 observing days). We obtain the following source size parameters for our best data set (2017): a major axis FWHM of \(146 \pm 12 \mu\text{as}\); a minor axis FWHM of \(99 \pm 7 \mu\text{as}\); and a position angle of \(70 \pm 6^\circ\), almost oriented along the diffractive kernel. This alignment may be coincidental, or it may indicate that the assumed diffractive kernel is incorrect, producing a biased intrinsic size that is aligned with the scattering kernel. Because our imaging of the 2017 data set in Issaoun et al. [2019b] yielded a largely unconstrained position angle, this alignment may also indicate that the Gaussian model for intrinsic structure is overly simplistic, with tight posteriors that are spurious. The derived major-to-minor axis ratio (or “axial ratio”) of \(1.5 \pm 0.2\) is slightly larger than the ratio \((1.2^{+0.3}_{-0.2})\) measured directly from the intrinsic image reconstructed by Issaoun et al. [2019b]. In addition to different systematics due to varying methods for the intrinsic size measurement, the \(\alpha\) and \(r_{in}\) parameters used to separate the scattering effects from the intrinsic source structure are different from the derived median values (although they are consistent within the 95% confidence interval for two of the three data sets), shown in Table 4.3. In Issaoun et al. [2019b] we assumed the J18 parameter values, whereas in this work we fit for \(\alpha\) and \(r_{in}\) within THEMIS simultaneously with the intrinsic source parameters. Thus, the degeneracy in the scattering and intrinsic source size parameters likely lead to this slight shift in the measurement for the axial ratio. Whereas in Issaoun et al. [2019b] we assumed particular values of \(\alpha\) and \(r_{in}\) to obtain information on the intrinsic source structure, our new result, however, confirms that the source still modestly deviates from circular symmetry even if we allow the scattering parameters to vary.

In Issaoun et al. [2019b] we explored a set of general-relativistic magnetohydrodynamic (GRMHD) simulations of the Sgr A* accretion flow: disk versus jet driven emission, varying particle acceleration, varying spin, and varying heating prescriptions [Mościbrodzka et al., 2009, 2014, 2016; Davelaar et al., 2018; Howes, 2010; Rowan et al., 2017; Chael et al., 2018a]. We found that high-inclination jet-dominated models (> \(20^\circ\) of face-on) produce larger axial ratios than the measurement on Sgr A*, and are thus ruled out. Our new axial ratio measurement in this work, allowing the freedom to fit the scattering parameters, confirms this result. This range in allowed inclination for jet models is also consistent with the independent near-infrared orbiting flare results from the Gravity experiment [Gravity Collaboration et al., 2018b]. In addition, the measured axial ratio is now slightly too large to also be consistent with low-inclination disk-dominated models (< \(20^\circ\)), which are close to circular [Figures 9 and 10 of Issaoun et al., 2019b]. With this new measurement, the explored emission models best able to replicate the size and axial ratio observed for Sgr A* are thus low-inclination jet or mid/high-inclination disk models. However it is worth noting that the degeneracy in the scattering and intrinsic source parameters could be the cause of this shift in the measured value of the axial ratio, and only a small subset of GRMHD models were studied. Therefore whether low-inclination disks are truly ruled out remains uncertain.

### 4.4.3 Constraints on the Power Spectrum

The refractive noise power on a long baseline \(u\) is chromatic, dependent on the power-law index \(\alpha\) of the phase fluctuations, and is dominated by refractive modes with \(q \sim 2\pi u/D\). Previous
Figure 4.10: Observational constraints, from refractive noise on long interferometric baselines, on the power spectrum $Q(q)$ of phase fluctuations. Note that $Q(q)$ is dimensionless and independent of the observing wavelength. Constraints at 3.6 and 1.3 cm are from J18; the 3.5 mm results are from the observations reported here and by Issaoun et al. [2019b]. The red diamonds represent constraints on the power for wavenumbers $q^{-1} \sim 10^{12} - 10^{14}$ cm from refractive noise on long baselines observed at 3.6 cm, 1.3 cm and new 3.5 mm wavelengths along the scattering major axis. The green shaded region delimits the range of modes that are expected to contribute refractive noise to 1.3 mm EHT images of Sgr A*. The power spectra of two scattering models are plotted: the GS06 model with $\alpha = 0$ and $r_{in} = 2 \times 10^6$ km (dashed gray); and the recommended J18 model with $\alpha = 1.38$ and $r_{in} = 800$ km (solid blue). All observational constraints are consistent with the recommended J18 model. While the previous cm wavelength detections of refractive noise are consistent with both models, the new 3.5 mm upper limit for the power from refractive noise along the scattering major axis is a factor of 10 below the predicted power by the GS06 model.
constraints on the power spectrum $Q(q)$ for a wavenumber $q^{-1}$ based on cm-wavelength refractive noise power detections by Johnson et al. [2018] showed that both the J18 and the GS06 models fit cm-wave constraints but the models are expected to be most different in the mm-wave regime. In our 2017 and 2018 data sets, we solely detect power on long baselines on ALMA baselines sampling mostly along the scattering minor axis. However, previous cm-wave measurements of noise power were along the scattering major axis [see Johnson et al., 2018, Figure 14]. To add a 3.5 mm constraint to those measurements, we therefore need to make use of estimated $4\sigma$ upper limits on baselines along the scattering major axis where no detections were found on Sgr A* but the level of noise for those baselines is known from detections on calibrators.

In Figure 4.10, we present a similar plot to Figure 14 in Johnson et al. [2018], with our newly added 3.5 mm constraint. Our mean visibility amplitude (after noise-debiasing) is 6 mJy along the scattering minor axis on baselines beyond 1.8 G\(\lambda\), measured on ALMA baselines for our 2017–2018 data sets. We determine the central value (red diamond) for the major axis constraint as the equivalent power measured along the scattering major axis at 3.5 mm for a scattering model giving 6 mJy of refractive noise along the minor axis. The upper limit of the constraint is given by the $4\sigma$ upper limit of 30 mJy at 1.8 G\(\lambda\) on the sensitive east-west GBT–IRAM 30 m baseline from our 2017 data set [Issaoun et al., 2019b]. The lower limit of the constraint assumes that the 6 mJy detection for a single scattering realization is purely refractive noise, giving a lower-limit of $\sim$3 mJy for the ensemble-average RMS [see Section 6.1 of Issaoun et al., 2019b]. Because we are combining estimates of refractive noise along the major and minor axes, this estimate of $Q(q)$ is sensitive to the assumed model of magnetic field wander [here, we use the “dipole” model of Psaltis et al., 2018] but primarily depends on the rough extent of the ensemble-average image so is insensitive to the assumed scattering parameters, $\alpha$ and $r_{in}$ [see, e.g., Eq. 16 of Johnson & Narayan, 2016].

We also plot in Figure 4.10 the power spectra for the two scattering models, GS06 (dashed gray line) and J18 (blue solid line). Refractive modes impacting horizon-scale reconstructions with the EHT are within the green shaded region. While both models fit constraints in the cm-wave regime, the 3.5 mm constraint discriminates between the two: the power predicted by the J18 model is consistent with the constraint, while the power predicted by GS06 model lies an order of magnitude above it. The 3.5 mm measurement directly probes refractive modes within the EHT’s field of view, for which scattering substructure can contaminate EHT horizon-scale images. Assuming the J18 model, our 3.5 mm result confirms expectations of low (of the order $\sim 1\%$ of total flux density) levels of refractive noise on long VLBI baselines observable with the EHT at 1.3 mm.

### 4.5 Summary

We have presented 2018 observations of Sgr A* using ALMA in concert with the GMVA at 86 GHz. In combination with observations carried out in 2017 presented in Issaoun et al. [2019b], we show that the ALMA–GBT baseline resolves persistent non-Gaussian morphology along the scattering minor axis. In addition, long-baseline detections to ALMA (1.8–2.4 G\(\lambda\)) exhibit characteristics of low-level refractive noise across both years of observations. Using the scattering model developed by Psaltis et al. [2018], we show that these long-baseline detections are consistent with the
scattering parameters estimated in Johnson et al. [2018], while they are at least five times weaker than the expected refractive noise for the parameters suggested by Goldreich & Sridhar [2006]. While a single realization of the refractive scattering would give values this low for approximately 4% of observations, the probability of seeing two independent observations at this level is less than 0.15%, firmly ruling out the Goldreich & Sridhar [2006] parameters.

We made use of the chromatic nature of the interstellar scattering to put stringent constraints on the intrinsic source extent at 86 GHz along the scattering minor axis. We combined 8, 22, and 43 GHz constraints on the scattering parameters presented in Johnson et al. [2018] with our 86 GHz persistent flux density excess on the ALMA–GBT baseline. We found that the ALMA–GBT baseline flux density excess predicts an intrinsic source extent along the scattering minor axis of $\sim 100 \mu$as for Sgr A* for the ranges of power spectrum power-law index $\alpha$ and inner scale of turbulence $r_{in}$ allowed by lower-frequency measurements from Johnson et al. [2018].

Direct modeling of our 86 GHz data with Themis [Broderick et al., 2020], fitting simultaneously the scattering and intrinsic source parameters, gave overlapping $\alpha - r_{in}$ parameter ranges with lower-frequency constraints and source size estimates consistent with the lower-frequency predictions. Source size estimates for all independent data sets are consistent within their uncertainties. The fitted source size parameters obtained for our best data set, on 3 April 2017, are: a major axis FWHM of $146 \pm 12 \mu$as; a minor axis FWHM of $99 \pm 7 \mu$as; and a position angle of $70 \pm 6^\circ$, almost oriented along the diffractive kernel. We obtain a major-to-minor axis ratio for the source of $1.5 \pm 0.2$. In the scope of the set of general-relativistic magnetohydrodynamic simulations explored in Issaoun et al. [2019b], high-inclination jet-dominated models produce larger axial ratios are ruled out, but the measured axial ratio is now too large to also be consistent with low-inclination disk-dominated models [Figures 9 and 10 of Issaoun et al., 2019b]. Assuming the fitted scattering parameters, the explored emission models best able to replicate the size and axial ratio observed for Sgr A* are thus low-inclination jet or mid/high-inclination disk models.

We have shown that 86 GHz measurements of the interstellar scattering are crucial for the discrimination of scattering models. In particular, our observations favor a dissipation scale of $\sim 10^3$ km for interstellar turbulence, which is consistent with a characteristic scale determined by the ion Larmor radius [Spangler & Gwinn, 1990]. Irrespective of the specific scattering properties, our results probe the scales of phase fluctuations that are comparable to the angular size of the black hole shadow of Sgr A*, giving model-independent insights into how scattering may affect images of Sgr A* produced with the Event Horizon Telescope. Future observations at 86 GHz with the newly upgraded 50-m surface of the LMT will enable probing of the structure along the scattering minor axis on a highly sensitive baseline with ALMA, providing a second anchor for deviations from Gaussian morphology and model constraints. GMVA bandwidth enhancements and station expansion would also enable higher fidelity imaging, enable higher sensitivity in the east–west to possibly detect refractive structure along the scattering major axis in the mm waveband, and possibly obtain the first polarization detections on VLBI baselines at 86 GHz, allowing us to further sharpen our view of the accretion flow of Sgr A*.

**Acknowledgements**

We thank our anonymous referee for a thorough review and helpful suggestions, which improved the quality and clarity of the manuscript. We thank Vincent Fish for his helpful comments and
careful review. This work is supported by the ERC Synergy Grant “BlackHoleCam: Imaging the Event Horizon of Black Holes”, Grant 610058. We thank the National Science Foundation (AST-1716536, AST-1440254, AST-1935980, AST-2034306) and the Gordon and Betty Moore Foundation (GBMF-5278) for financial support of this work. This work was supported in part by the Black Hole Initiative, which is funded by grants from the John Templeton Foundation and the Gordon and Betty Moore Foundation to Harvard University. A.C. is supported by 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. I.C. is supported by the National Research Foundation of Korea (NRF) via a Global PhD Fellowship Grant (NRF-2015H1A2A1033752). L.L. acknowledges the financial support of DGAPA, UNAM (projects IN112417 and IN112820), and CONACyT, México (projects 275201 – Agencia Espacial Mexicana; and 263356 – Ciencia de Frontera). M.K. was supported by JSPS KAKENHI grant Nos. JP18K03656 and JP18H03721. This paper makes use of the following ALMA data: ADS/JAO.ALMA2017.1.00795.V. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The ALMA data required non-standard processing by the VLBI QA2 team (C. Goddi, I. Martí-Vidal, G. B. Crew, H. Rottmann, and H. Messias). This research has made use of data obtained with the Global Millimeter VLBI Array (GMVA), coordinated by the VLBI group at the Max-Planck-Institut für Radioastronomie (MPIfR). The GMVA consists of telescopes operated by MPIfR, IRAM, Onsala, Metsahovi, Yebes, the Korean VLBI Network, the Green Bank Observatory and the Very Long Baseline Array (VLBA). The VLBA is a facility of the National Science Foundation under cooperative agreement by Associated Universities, Inc. The data were correlated at the DiFX correlator of the MPIfR in Bonn, Germany. This work made use of the Swinburne University of Technology software correlator [Deller et al., 2011], developed as part of the Australian Major National Research Facilities Programme and operated under licence.
First M87 Event Horizon Telescope Results III. Data Processing and Calibration

The Event Horizon Telescope Collaboration et al.

Abstract

We present the calibration and reduction of Event Horizon Telescope (EHT) 1.3 mm radio wavelength observations of the supermassive black hole candidate at the center of the radio galaxy M87 and the quasar 3C279, taken during the 2017 April 5–11 observing campaign. These global very-long-baseline interferometric observations include for the first time the highly sensitive Atacama Large Millimeter/submillimeter Array (ALMA); reaching an angular resolution of 25 μas, with characteristic sensitivity limits of ~1 mJy on baselines to ALMA and ~10 mJy on other baselines. The observations present challenges for existing data processing tools, arising from the rapid atmospheric phase fluctuations, wide recording bandwidth, and highly heterogeneous array. In response, we developed three independent pipelines for phase calibration and fringe detection, each tailored to the specific needs of the EHT. The final data products include calibrated total intensity amplitude and phase information. They are validated through a series of quality assurance tests that show consistency across pipelines and set limits on baseline systematic errors of 2% in amplitude and 1° in phase. The M87 data reveal the presence of two nulls in correlated flux density at ~3.4 and ~8.3 Gλ and temporal evolution in closure quantities, indicating intrinsic variability of compact structure on a timescale of days, or several light-crossing times for a few billion solar-mass black hole. These measurements provide the first opportunity to image horizon-scale structure in M87.
5.1 Introduction

The principle of very-long-baseline interferometry (VLBI) is to connect distant radio telescopes to create a single virtual telescope. On the ground, VLBI enables baseline lengths comparable to the size of the Earth. This significantly boosts angular resolution, at the expense of having a non-uniform filling of the aperture. In order to reconstruct the brightness distribution of an observed source, VLBI requires cross-correlation between the individual signals recorded independently at each station, brought to a common time reference using local atomic clocks paired with the Global Positioning System (GPS) for coarse synchronization. The resulting complex correlation coefficients need to be calibrated for residual clock and phase errors, and then scaled to physical flux density units using time-dependent and station-specific sensitivity estimates. Once this process is completed, further analysis in the image domain can refine the calibration using model-dependent self-calibration techniques [e.g., Pearson & Readhead, 1984; Wilkinson, 1989]. For more details on the principles of VLBI, see, e.g., Thompson et al. [2017].

At centimeter wavelengths, the technique of VLBI is well established. Correlation and calibration have been optimized over decades, resulting in standard procedures for the processing of data obtained at national and international facility instruments, such as the Very Long Baseline Array\(^1\) (VLBA), the Australian Long Baseline Array\(^2\) (LBA), the East Asian VLBI Network\(^3\) (EAVN), and the European VLBI Network\(^4\) (EVN). At higher frequencies, the increased effects from atmospheric opacity and turbulence pose major challenges. The characteristic atmospheric coherence timescale is only a few seconds for millimeter wavelengths, and sensitivity must be sufficient to track phase variation over correspondingly short timescales. Large collecting areas and wide bandwidths prove essential when observing even the brightest continuum sources over a range of elevations and reasonable weather conditions. Furthermore, the transfer of phase solutions from a bright calibrator to a weak source, typically done at centimeter wavelengths, is not feasible at high frequencies, because differential atmospheric propagation effects are more significant, and because there are few bright, compact calibrators.

The Event Horizon Telescope (EHT) is a global VLBI array of millimeter and submillimeter wavelength observatories with the primary goal of studying the strong gravity, near–horizon environments of the supermassive black holes in the Galactic Center, Sagittarius A\(^*\) (Sgr A\(^*\)), and at the center of the nearby radio galaxy M87 [Doeleman et al., 2009a; Event Horizon Telescope Collaboration et al., 2019b, hereafter Paper II]. In April 2017, the EHT conducted science observations at a wavelength of \(\lambda \simeq 1.3\) mm, corresponding to a frequency of \(\nu \simeq 230\) GHz. The network was joined for the first time by the Atacama Large Millimeter/submillimeter Array (ALMA) configured as a phased array, a capability developed by the ALMA Phasing Project [APP, Doeleman, 2010; Fish et al., 2013; Matthews et al., 2018]. The addition of ALMA, as a highly sensitive central anchor station, drastically changes the overall characteristics and sensitivity limits of the global array [Paper II].

Although operating as a single instrument spanning the globe, the EHT remains a mixture of new and well-exercised stations, single-dish telescopes and phased arrays with varying designs

\(^1\)https://science.nrao.edu/facilities/vlba.
\(^3\)https://radio.kasi.re.kr/eavn.
\(^4\)http://www.evlbi.org.
and operations. Each observing cycle over the last several years has been accompanied by the introduction of new telescopes to the array, and/or significant changes and upgrades to existing stations, data acquisition hardware, and recorded bandwidth [Paper II]. EHT observations result in data spanning a wide range of signal-to-noise ratio (S/N) due to the heterogeneous nature of the array, and the high observing frequency produces data that are particularly sensitive to systematics in the signal chain. These factors, along with the typical challenges associated with VLBI, have motivated the development of specialized processing and calibration techniques.

In this paper we describe the full data processing pathway and pipeline convergence leading to the first science release (SR1) of the EHT 2017 data. Given the uniqueness of the data set and scientific goal of the EHT observations, our processing focuses on the use of unbiased automated procedures, reproducibility, and extensive review and cross-validation. In particular, data reduction is carried out with three independent phase calibration (fringe fitting) and reduction pipelines. The Haystack Observatory Processing System [HOPS; Whitney et al., 2004] has been the standard for calibrating EHT data from prior observations [e.g., Doeleman et al., 2008; Fish et al., 2011; Doeleman et al., 2012; Johnson et al., 2015; Akiyama et al., 2015; Fish et al., 2016; Lu et al., 2018]. HOPS reduction of the 2017 data is supported by a suite of auxiliary calibration scripts to form the EHT-HOPS pipeline [Blackburn et al., 2019]. The Common Astronomy Software Applications package [CASA; McMullin et al., 2007] is primarily aimed at processing connected-element interferometer data. The recent addition of a fringe fitter and reduction pipeline has enabled the use of CASA for high frequency VLBI data processing [Janssen et al., 2019b, I. van Bemmel et al. 2019, in preparation]. The NRAO Astronomical Image Processing System [AIPS; Greisen, 2003] is the most commonly used reduction package for centimeter VLBI data. For this work, an automated ParselTongue [Kettenis et al., 2006] pipeline was constructed and tailored to the needs of EHT data reduction in AIPS.

The SR1 data consist of Stokes I complex interferometric visibilities of M87 and the quasar 3C279, corresponding to spatial frequencies of the sky brightness distribution sampled by the interferometer. M87 data indicate the presence of a resolved compact emission structure on a spatial scale of a few tens of µas, persistent throughout the week-long observing campaign. Closure phases and closure amplitudes unambiguously reflect non-trivial brightness distributions on M87 for the first time. They display broad consistency over different days, and in certain cases show clear evolution. A detailed analysis of this near-horizon scale structure is the subject of companion papers [Event Horizon Telescope Collaboration et al., 2019a,d,e,f, hereafter Papers I, IV, V and VI].

This paper is organized as follows. Section 5.2 presents an overview of the April 2017 observations. In Section 5.3 we outline the data flow from observations to science-ready data sets. We describe the correlation process in Section 5.4, the phase calibration process via three independent fringe fitting pipelines in Section 5.5, and the common flux density calibration scheme and amplitude error budget in Section 5.6. We give an overview of SR1 data products and a rudimentary description of their most evident, remarkable properties in Section 5.7. We present data set validation procedures and tests, estimates of systematic errors, and inter-pipeline comparisons in Section 5.8. Conclusions are given in Section 2.7.
Chapter 5: EHT Data Processing and Calibration

5.2 Observations

The EHT 2017 science observing run was scheduled for five nights during the ten-night 2017 April 5–14 (UTC) window with eight participating observatories at six distinct geographical locations, shown in Figure 5.1: 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, and the Submillimeter Array (SMA) and the James Clerk Maxwell Telescope (JCMT) on Mauna Kea in Hawaii. A detailed description of the EHT array is presented in Paper II. The 2017 science observing run consisted of observations of six science targets: the primary EHT targets SagittariusA* and M87, and the secondary targets 3C279, OJ287, Centaurus A, and NGC1052.

An array-wide go/no-go decision was made a few hours before the start of each night’s schedule, based on weather conditions and technical readiness at each of the participating observatories. A dry run of the go/no-go decision making was performed on April 4 to assess triggering and readiness procedures. All sites were technically ready and with good weather on the first night of the observing window. Observations were triggered on 2017 April 5, 6, 7, 10, and 11. Table 5.1 shows the median zenith sky opacities for each of the triggered days. April 8 was not triggered due to thunderstorms at the LMT, SMT shutdown due to strong winds and the need to run technical tests at ALMA. April 9 was not triggered due to a chance of the SMT remaining closed.
5.2 Observations

Table 5.1: Median zenith sky opacities (1.3 mm) at EHT sites during the April 2017 observations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Median Zenith $\tau_{1.3\text{mm}}$</th>
<th>Apr 5</th>
<th>Apr 6</th>
<th>Apr 7</th>
<th>Apr 10</th>
<th>Apr 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA/APEX</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>SMA/JCMT</td>
<td>0.10</td>
<td>0.07</td>
<td>0.09</td>
<td>0.05</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>0.18</td>
<td>0.13</td>
<td>0.14</td>
<td>0.10</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>LMT</td>
<td>0.13</td>
<td>0.16</td>
<td>0.21</td>
<td>0.26</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>0.21</td>
<td>0.28</td>
<td>0.23</td>
<td>0.19</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>

Note – Median zenith sky opacities are measured at each site and reported through station log files and the VLBI monitor as described in Paper II.

due to strong winds and LMT snow forecast. Weather was good to excellent for all other stations throughout the observing window.

In addition to favorable weather conditions, operations at all sites were successful and resulted in fringe detections across the entire array. A number of mild to moderate site and data issues were uncovered during the analysis, and their detailed characterization and mitigation are given in Appendix 5.A. Notable issues affecting processing, calibration, and data interpretation are: (1) a clock frequency instability at PV resulting in $\sim 50\%$ amplitude loss to that station, (2) recorder configuration issues at APEX resulting in a significant number of data gaps and low data validity at correlation, (3) pointing errors at LMT, large compared to the beam, resulting in unpredictable amplitude loss and inter and intra-scan gain variability, and (4) a common local oscillator used at SMA and JCMT resulting in opposite sideband contamination at the level of $\sim 15\%$ for short integration times, making the SMA–JCMT intra-site baseline less useful for calibration. All known issues with a significant effect on the data are addressed at various stages of processing and calibration, although some (such as residual gains at the LMT, and SMA–JCMT sideband contamination) necessitate additional care taken during data interpretation.

M87 ($\alpha_{2000} = 12^h30^m49^s.42, \delta_{2000} = 12^\circ23'28''0.42$) was observed as a target source on three nights (2017 April 5, 6, and 11). In addition, seven scans on M87 were included as a calibration source (for 3C279) on 2017 April 10. Each of the four tracks consists of multiple scans lasting between 3 and 7 minutes. In most tracks, VLBI scans on M87 began when it rose at the LMT and ended when it set below $20^\circ$ elevation at ALMA. Scans on M87 were interleaved with scans on the quasar 3C279 ($\alpha_{2000} = 12^h56^m11^s.17, \delta_{2000} = -05^\circ47'21''52$), another EHT target with a similar right ascension. The observed schedules for M87 and 3C279 during the 2017 campaign are shown in Figure 5.2. The schedules were optimized for wide $(u, v)$ coverage on all target sources when possible. All stations apart from the JCMT observed with full polarization. The JCMT observed a single circular polarization component per night (right circular polarization, RCP, for April 5 and 6, left circular polarization, LCP, for April 10 and 11).

The 2017 observing run recorded two 2 GHz bands, low and high band, centered at sky frequencies of 227.1 and 229.1 GHz respectively, onto Mark6 VLBI recorders [Whitney et al., 2004] at an aggregate recording rate of 32 Gbps with 2-bit sampling. All telescopes apart from ALMA
observed in circular polarization with the installation of quarter-wave plates. Single-dish sites used block-down converters (BDCs) to convert the intermediate frequency (IF) signal from the front-ends to a common 0–2 GHz baseband, which was digitally sampled via Reconfigurable Open Architecture Computing Hardware 2 (ROACH2) digital backends [R2DBEs; Vertatschitsch et al., 2015]. The SMA observed as a phased array of 6 or 7 antennas for which the phased sum signal was processed in the SWARM correlator [see Primiani et al., 2016; Young et al., 2016, for more details]. ALMA observed as a phased array of 37 dual linear polarization antennas for which the phased sum signal was processed in the Phasing Interface Cards (PICs) installed at the ALMA baseline correlator [see Matthews et al., 2018, for more details]. Instrumentation development leading up to the 2017 observations is presented in Paper II.

5.3 Data Flow

The EHT data flow from recording to analysis is outlined in Figure 5.3. Through the receiver and backend electronics at each telescope, the sky signal is mixed to baseband, digitized and recorded directly to hard disk, resulting in petabytes of raw VLBI voltage signal data. The correlator uses an a priori Earth geometry and clock/delay model to align the signals from each telescope to a common time reference, and estimates the pair-wise complex correlation coefficient \( r_{ij} \) between antennas. For signals \( x_i \) and \( x_j \) between stations \( i \) and \( j \),

\[
 r_{ij} = \frac{\langle x_i x_j^\ast \rangle}{\eta_Q \sqrt{\langle x_i x_i^\ast \rangle \langle x_j x_j^\ast \rangle}}, \tag{5.1}
\]

where \( \eta_Q \) represents a digital correction factor to compensate for the effects of low-bit quantization. For optimal 2-bit quantization, \( \eta_Q \approx 0.88 \).

---

*Figure 5.2: EHT 2017 observing schedules for M87 and 3C 279 covering the four days of observations. Empty rectangles represent scans that were scheduled, but were not observed successfully due to weather, insufficient sensitivity, or technical issues. The filled rectangles represent scans corresponding to detections available in the final data set. Scan duration varies between 3–7 minutes, as reflected by the width of each rectangle.*
Figure 5.3: Data processing pathway of an EHT observation from recording to source parameter estimation (images, or other physical parameters). At the calibration stage, instrumental and environmental gain systematics are estimated and removed from the data so that a smaller and simpler data product can be used for source model fitting at a downstream analysis stage.

The correlation coefficient may vary with both time and frequency. For FX correlators, signals from each antenna are first taken to the frequency domain using temporal Fourier transforms on short segments (F), and then pair-wise correlated (X). The expectation values in Equation 5.1 are calculated by averaging over time-frequency volumes where the inner products remain stable. At millimeter wavelengths, a correlator can average around 1 second $\times$ 1 MHz, or $2 \times 10^6$ samples, before clock errors such as residual delay, delay-rate (e.g., Doppler shift), and stochastic changes in atmospheric path length cause unwanted decoherence in the signal (Section 5.4). The post-correlation data reduction pipeline models and fits these residual clock systematics, allowing data to be further averaged by a factor of $10^3$ or more, to the limits imposed by intrinsic source structure and variability (Section 5.5). For many EHT baselines, the astronomical signal is not detectable above the noise until phase corrections resulting from these calibration solutions are applied and the data are coherently (vector) averaged.

In addition to reducing the overall volume and complexity of the data, the calibration process attempts to relate the pairwise correlation coefficients $r_{ij}$, which are in units of thermal noise of the detector, to correlated flux density in units of Jansky (Jy),

$$r_{ij} = \gamma_i \gamma_j^* V_{ij}. \quad (5.2)$$

The visibility function, $V_{ij}$, represents the mutual coherence of the electric field between ends of
the baseline vector joining the sites, projected onto the plane of propagation. For an ideal interferometer, \( V_{ij} \) samples a Fourier component of the brightness distribution on the sky [via the van Cittert-Zernike theorem, van Cittert, 1934; Zernike, 1938; Thompson et al., 2017]. The dimensionless spatial frequency \( u = (u, v) \) of the Fourier component is determined by the projected baseline expressed in units of the observing wavelength. Here, we have made the implicit assumption that the relationship between correlation coefficient and visibility can be factored into complex station-based forward gains \( \gamma_i \) and \( \gamma_j \). This process of flux density calibration requires an a priori assessment of the sensitivity of each antenna in the array, captured by the system-equivalent flux density \( \text{SEFD}_i = |1/\gamma_i|^2 \) of the thermal noise power, as described in Section 5.6.

After the basic calibration and reduction process, the data are passed through additional postprocessing tasks to further average the data to a manageable size for source imaging and model fitting, and to apply any network self-calibration constraints based on independent a priori assumptions about the source, such as large-scale (milliarcsecond and larger) structure, total flux density, and degree of total polarization (Section 5.6.2). The final network-calibrated data products are further averaged to a 10 second segmentation in time and across each 2 GHz band to provide smaller files for downstream analysis (Section 5.7.1).

### 5.4 Correlation

The recorded data from each station were split by frequency band and sent to MIT Haystack Observatory and the Max-Planck-Institut für Radioastronomie (MPIfR) for correlation, as described in Paper II. The Haystack correlator handled the low frequency band (centered at 227.1GHz), with MPIfR correlating the high band (centered at 229.1GHz). Each correlator is a networked computer cluster running a standard installation of the DiFX software package [Deller et al., 2011].

The correlators use a model (calc11) of the expected wavefront arrival delay as a function of time on each baseline. The delay model very precisely takes into account the geometry of the observing array at the time of observation, the direction of the source, and a model of atmospheric delay contributions [e.g., Romney, 1995]. Baseband data on a few high-S/N scans with good coverage were exchanged between the two sites to verify the output of each correlator against the other.

Data were correlated with an accumulation period (AP) of 0.4s and a frequency resolution of 0.5 MHz (Figure 5.4). Due to the need to rationalize frequency channelization with the ALMA setup [each 1.875 GHz spectral window at ALMA is broken up into 32 spectral IFs of 62.5 MHz, separated by 58.59375 MHz and thus slightly overlapping; Matthews et al., 2018], the frequency points are grouped into IFs that are 58 MHz wide (using DiFX zoom mode), each with 116 individual channels and a small amount of bandwidth discarded between spectral IFs.

At the SMA, the original data are recorded in the frequency domain rather than the time domain, owing to the architecture of the SMA correlator. Moreover, the recorded frequency range of 2288 MHz is slightly larger and offset by 150 MHz from the frequency range at the other non-ALMA sites. An offline preprocessing pipeline, called APHIDS [Primiani et al., 2016], is used to perform the necessary filtering, frequency conversion, and transformation to the time domain, so that the format of the SMA data delivered to the VLBI correlator is the same as for single-dish stations. Part of the necessary offline preprocessing includes deriving clock offsets on a scan-by-scan basis for the delivered data. These offsets are determined by cross-correlating the
5.4 Correlation

**Figure 5.4:** Time and frequency resolution of EHT 2017 data as it is recorded and processed. Correlation parameters for the EHT are chosen to be compatible with ALMA’s recorded sub-bands that are 62.5 MHz wide, overlap slightly, and have starting frequencies aligned to $1/(32 \mu s)$. The raw output after calibration and reduction maintains the original correlator accumulation of 0.4s, but averages over each 58 MHz spectral IF, centered on each ALMA sub-band. The data are further averaged at the network amplitude self-calibration stage (not shown) for a more manageable data volume.

Preprocessed SMA data with separate data recorded with an R2DBE-Mark6 pair, taking a 2nd IF signal from the SMA reference antenna as input.

The IF from the JCMT was recorded using backend equipment installed at the SMA [Paper II]. This was achieved by transporting the first IF from the JCMT to the SMA, where the second downconversion, digitization and recording were done. Since the second downconversion at the SMA introduces a net offset of 150 MHz with respect to the nominal EHT RF band, this means that the recorded JCMT data sent to the correlator are subject to the same frequency offset. The mismatch eliminates one of the thirty-two 58 MHz spectral IFs in the final correlation for JCMT baselines.

ALMA observes linear polarization, while the rest of the EHT observes circular polarization. The software routine PolConvert [Martí-Vidal et al., 2016b; Matthews et al., 2018] was created to convert visibilities, output from the correlator in a mixed-polarization basis, to the pure circular basis of the EHT. PolConvert takes auxiliary calibration input from the quality assurance stage 2 (QA2) ALMA interferometric reduction of data [Goddi et al., 2019]. Execution of the PolConvert tool completes the correlation (circularized visibilities on baselines to ALMA) and provides final ANTAB\textsuperscript{6} format data for flux density calibration of the ALMA phased array. The original native (Swinburne format) correlator output from DiFX is converted using available DiFX tools to a Mark4 [Whitney et al., 2004] compatible file format for processing through HOPS, and to FITS-IDI [Greisen, 2011] files for further processing with AIPS and CASA.

\textsuperscript{6}Free-format parsable text file containing flux density calibration information and keywords as defined for AIPS: http://www.aips.nrao.edu/cgi-bin/ZXHLP2.PL?ANTAB.
5.5 Fringe Detection

In the limit for which all correlator delay model parameters were known perfectly ahead of time and there were no atmospheric variations, the model delays would exactly compensate for the delay on each baseline of the data, and the correlated data could be coherently integrated in time and frequency to build up sensitivity. In practice, many of the model parameters are not known exactly at correlation. For example, the observed source may have structure and may be centered at an offset from the expected coordinates, the position of each telescope may differ from the best estimate, instrumental electronic delays may not be known, or variable water content in the atmosphere may cause the atmospheric delay to deviate from the simple model. It is therefore necessary to search in delay and delay-rate space for small corrections to the model values that maximize the fringe amplitude: this process is known in VLBI data processing as fringe fitting [e.g., Cotton, 1995]. In this section, we describe three independent fringe fitting pipelines for phase calibration, based on three different software packages for VLBI data processing: HOPS (Section 5.5.1), CASA (Section 5.5.2), and AIPS (Section 5.5.3).

5.5.1 HOPS Pipeline

The Haystack Observatory Postprocessing System (HOPS) is a collection of software packages and data framework designed to analyze and reduce output from a Mark III, IV, or DiFX correlator. It has been used extensively for the processing of early EHT data [Doeleman et al., 2008; Fish et al., 2011; Doeleman et al., 2012; Johnson et al., 2015; Akiyama et al., 2015; Fish et al., 2016; Lu et al., 2018]. For EHT 2017 observations, HOPS was augmented with a collection of auxiliary calibration scripts, and packaged into an EHT-HOPS pipeline [Blackburn et al., 2019] for automated processing of this and similar data sets. Compared to the reduction of data from previous runs, the EHT-HOPS pipeline is unique in that it finds a single self-consistent global fringe solution (station-based delays, delay-rates, and instrumental and atmospheric phase) for calibration. The pipeline also provides standard UVFITS formatted visibility data products for downstream analysis.

The EHT-HOPS pipeline processes output from the DiFX correlator that has been converted to Mark4 format via the DiFX tool difx2mark4. This conversion process includes normalization by auto-correlation power per 58 MHz spectral IF in each AP of 0.4 s (Figure 5.4), as well as a 1/0.88252 amplitude correction factor for 2-bit quantization efficiency. Stages of the pipeline (Figure 5.5) run the HOPS fringe fitter fourfit several times (once per stage) while making iterative corrections to the phase calibration applied to the data before solving for delays and delay-rates. The initial setup (default config, flags – Figure 5.5) includes manual flagging (removal of bad data) in time and frequency, as well as an ALMA-specific correction for digital phase offsets between spectral IFs.

ALMA is used as a reference station for estimating stable instrumental phase (phase bandpass) and relative delay between right and left circular polarization (R-L delay offsets) for remote stations. The estimates are done using S/N weighted averages of the strong ALMA baseline measurements. Here we make use of the fact that ALMA RCP and LCP data are already delay

---

7 https://www.haystack.mit.edu/tech/vlbi/hops.html
Figure 5.5: Stages of the EHT-HOPS pipeline and postprocessing steps, as described in the text. The first five stages, shown in the left box, are iterations of HOPS fringe fitter *fourfit*. Here a comprehensive phase calibration model is gradually built for the data. At the end of the five *fourfit* stages, the correlation coefficients are evaluated at a single global (station-based) set of relative delays and delay-rates. The data are then converted to *UVFITS* format, and a remaining suite of postprocessing tools provide amplitude calibration and time-and-polarization dependent phase calibration.
and phase calibrated during the QA2/PolConvert process [Goddi et al., 2019]. For rapid non-linear phase (atmospheric phase) that varies over seconds and that must be calibrated on-source, the strongest station (generally ALMA when it is present; see also Section 2 of Paper II) is automatically determined for each scan based on signal-to-noise, and is used as a phase reference. Baselines to the reference station are then used to phase stabilize the remaining sites.

Due to the large number of free parameters involved in correcting for atmospheric phase, a leave-out-one cross-estimation approach is adopted for this step to avoid self-tuning. For each baseline, a smooth phase model is estimated by stacking RCP and LCP data over 31 (of 32) spectral IFs. The estimated phase from the 31-IF average is used to correct the remaining IF, and the process cycles through IFs to cover the full band. In this way, phase corrections are never estimated from the same data they are applied to, which avoids introducing false coherence from self-tuning to random thermal noise and introducing a positive bias to amplitudes. The effective solution interval for the phase model depends on S/N, and is chosen per baseline to balance anticipated atmospheric phase drift with thermal noise in the estimate. Additional a priori corrections for small residual clock frequency offsets after correlation (Appendix 5.A) are made here as well.

During a final reduction with fourfit (close fringe solution), rather than fitting for unconstrained delays and delay-rates per baseline and polarization product, a single set of station-based delays and delay-rates is fixed corresponding to a global fringe solution. These are derived from a least-squares solution [as proposed by Alef & Porcas, 1986] to relative delays and delay-rates from confident baseline detections with S/N > 7, and stations which remain unconstrained by this process are removed from the data set. No interpolation of these fringe solutions is performed across sources and scans; instead precise closure of delay and delay-rate from strong baseline detections is required to report any measurement on a weak baseline. Correlation coefficients on baselines with no detectable signal are still calculated (Figure 5.11, where S/N < few), but only when the relative clock model is constrained through other baseline detections.

The resulting complex visibility data are converted to UVFITS format, and amplitude calibration is done in the EHT Analysis Toolkit’s (eat) postprocessing framework, shared by all pipelines and described in Section 5.6. For the HOPS pipeline, calibration of complex polarization gain ratios is performed in a postprocessing stage rather than during fourfit. Deterministic field rotation from parallactic angle and receiver mount type is corrected as a complex polarization-dependent a priori gain factor, and a smoothly-varying polynomial model is fit over many sources and used to correct residual RCP–LCP phase drift for each station. Details for all steps can be found in Blackburn et al. [2019].

The EHT-HOPS pipeline was additionally used for the reduction of observations of Sgr A* and calibrators at 86 GHz, with the Global Millimeter VLBI Array (GMVA) joined by ALMA. Despite the magnitude difference in bandwidth, a similar reduction to EHT data was performed on the GMVA data set. ALMA baselines were used to estimate stable instrumental phase and delay corrections. Baselines to either ALMA or the Green Bank Telescope (GBT) were used, due to their high S/N, to correct for stochastic atmospheric phase fluctuations on timescales of a few seconds. The performance of the pipeline on the GMVA data is described in Blackburn et al. [2019].

---

8http://github.com/sao-eht/eat
9https://www3.mpifr-bonn.mpg.de/div/vlbi/globalmm
et al. [2019], while scientific results from the data set are validated against historical observations in Issaoun et al. [2019b] (Chapter 3).

### 5.5.2 CASA Pipeline

The Common Astronomy Software Applications [CASA: McMullin et al., 2007] package was developed by NRAO to process data acquired with the JVLA and ALMA connected-element interferometers and in recent years has become the standard software for the calibration and analysis of radio-interferometric data. A newly developed fringe fitting task `fringefit` (I. van Bemmel et al. 2019, in preparation) has added the necessary delay and delay-rate calibration capabilities for VLBI. The modular, general-purpose `rPICARD` VLBI data reduction pipeline [Janssen et al., 2019b] is used for the calibration of EHT data. This section describes the incremental `rPICARD` calibration steps for EHT data, summarized in Figure 5.6.

The `importfitsidi` CASA task is used to import the FITS-IDI correlator output into CASA. Additionally, a digital correction factor for the 2-bit recorder sampling is applied when the data are loaded. Bad data are flagged based on text files compiled from station logs and known sources of RFI in stations’ signal chains with the `flagdata` task before performing the incremental calibration procedures. The `accor` task is used to scale the auto-correlations to unity and adjust the cross-correlations accordingly, correcting for incorrect sampler settings from the data recording stage. This is done for each 58MHz spectral IF individually, thereby correcting for a coarse bandpass at each station. This amplitude bandpass is refined by dividing the data by the auto-correlations at the 0.5 MHz channel resolution.

The phase calibration is done with the `fringefit` task, which solves for station-based residual post-correlation phases, delays, and rates with respect to a chosen reference station [Schwab & Cotton, 1983]. Unlike the HOPS pipeline, where field rotation angles are corrected a posteriori, `rPICARD` applies field rotation angle gain solutions on-the-fly, i.e., before each phase calibration correction. The most sensitive station is picked as reference in each scan. Eventually, all fringe solutions are re-referenced with the CASA `rerefant` task to a common station for each observing track to ensure phase continuity across scans.

Phases are first calibrated for the high S/N calibrator sources, which are used to correct for instrumental effects. Optimal time solution intervals to calibrate atmospheric intra-scan phase fluctuations ($\tau_{\text{sol}}$) are determined automatically based on the S/N of the data. The search is done for short solution intervals, close to the coherence time, which still yield detections on all possible baselines [Janssen et al., 2019b]. Typical solution intervals range from 2 to 30 s. Using these solution intervals, phases and rates are calibrated to extend the coherence time of the calibrator scans. This results in high S/N scan-based fringe solutions per 58 MHz spectral IF, which are used to obtain calibration solutions for instrumental effects. ALMA-induced phase offsets between spectral IFs are corrected with the short ALMA–APEX baseline. All baselines in the array are used by the global fringe fitter in the next step to solve for residual instrumental phase and delay offsets for all stations. After removing these instrumental data corruptions, a final `fringefit` step solves for multi-band delays on the (previously-determined) solution intervals. A 60-second median window filter is used to smooth the slowly varying multi-band delays, which effectively removes potential outliers. After fringe fitting, the phases are coherent in time and frequency, and
Figure 5.6: EHT data processing stages of rPICARD. Instrumental amplitude calibration effects are described in the top left box. Phases for the calibrator sources are corrected first to solve for instrumental effects (second box) and science targets are phase calibrated after the instrumental effects have been solved (third box). Finally, postprocessing steps are done outside of CASA for amplitude calibration (fourth box).
the **bandpass** task is used to solve for the frequency-dependent phase gains within each 58 MHz spectral IF for each station, using the combined data of all calibrator sources.

After all instrumental effects are calibrated out, the optimal fringe fit solution intervals \( T_{\text{sol}} \) are determined for the weaker science targets and phases, delays, and rates are solved for in a single **fringefit** step. The intra-scan fringe fritting on short solution intervals flags low S/N segments where no fringes are found to a specific station, e.g. when a station arrived late on source. Finally, the **exportuvfits** task is used to export the calibrated data from internal *Measurement Set* format to *UVFITS* files, which are then flux-density and network-calibrated in the common postprocessing framework.

Janssen et al. [2019b] demonstrate the **rPICARD** calibration capabilities in a close comparison with a traditional AIPS-based calibration using 43 GHz VLBA data of M87. The resultant image of the jet and counter-jet, which reveals a complex collimation profile, is in good agreement with earlier results from the literature [e.g., Walker et al., 2018]. The **rPICARD** pipeline was further used for the generation of synthetic EHT data [Paper IV], where known input delay and phase offsets were recovered as a ground-truth validation.

### 5.5.3 AIPS Pipeline

The Astronomical Image Processing System [AIPS; Greisen, 2003] is the most widely used software package for VLBI data reduction and processing at frequencies at or below \( \sim 86 \) GHz. It is commonly used in the VLBI community and was built to process low-S/N data from fairly homogeneous centimeter-wave observatories at low recording bandwidths. The EHT, however, falls in a different category: its high recording bandwidth and heterogeneous array produce data with a wide range of S/N, often dominated by systematic effects instead of thermal noise. These properties required the development of a custom pipeline based on AIPS, deviating from standard fringe fitting procedures for lower frequency data processing as outlined in e.g., the AIPS Cookbook.\(^{10}\)

The custom AIPS pipeline is an automated **Python**-based script using functions implemented in the **eat** package. It makes use of ** ParselTongue** [Kettenis et al., 2006], which provides a platform to manipulate AIPS tasks and data outside of the AIPS interface. The pipeline is summarized in Figure 5.7 and shows individual tasks used for calibration. A suite of diagnostic plots, using tasks **VPLOT** and **POSSM**, are also generated at each calibration step within the pipeline.

The loading of EHT data into AIPS, during which digital corrections for 2-bit quantization efficiency are applied, requires a concatenation of several packaged **FITS-IDI** files and a careful handling of the JCMT, which observes with a slightly shifted IF setup of the band (Section 5.4). The pipeline reduces each band (low and high) in separate runs. Data inspection and flagging of spurs in the frequency domain from accumulated scalar bandpass tables (generated with **BPASS**) and drop-outs or amplitude jumps in the time domain are done interactively with the AIPS tasks **BPEDT** and **EDITA**. The flags are saved in output flag tables to use in non-interactive reruns of the pipeline. Standard amplitude normalization steps are performed with the AIPS task **ACSCL**. The field rotation angle corrections are performed with an EHT-specific receiver mount correction

\(^{10}\) [http://www.aips.nrao.edu/cook.html.](http://www.aips.nrao.edu/cook.html)
Figure 5.7: Stages of the AIPS fringe fitting pipeline and postprocessing steps. The pipeline begins with direct data editing (interactively or via input correction and flag tables) and amplitude normalization (first box). The phase calibration process then follows via four steps with the AIPS fringe fitter `KRING` to solve for phase and delay offsets and rates (second box). Finally, postprocessing steps are done outside of AIPS for amplitude calibration (third box).
5.6 Flux Density Calibration

The flux density calibration for the EHT is done in two steps and is a common postprocessing procedure for all three phase calibration pipelines, as it involves very little handling of the data themselves. In Section 5.6.1, we describe the a priori calibration process to calibrate visibility amplitudes to a common flux density scale across the array. In Section 5.6.2, we present the network calibration process, where we use array redundancy to absolutely calibrate stations with an intra-site companion.

5.6.1 A Priori Amplitude Calibration

A priori amplitude calibration serves to calibrate visibility amplitudes from correlation coefficients to flux density measurements, as in Equation 5.2. Since the normalized correlation coefficients are in units of noise power, it is necessary to account for telescope sensitivities to convert to a uniform flux density scale across the array. The system-equivalent flux density (SEFD) of a radio telescope is the total system noise represented in units of equivalent incident flux density above the atmosphere. It can be written as

\[
\text{SEFD} = \frac{T_{\text{sys}}}{\text{DPFU}} \times \eta_{\text{el}},
\]

using the three measurable parameters,

1. \(T_{\text{sys}}\): the effective system noise temperature describes the total noise characterization of the system corrected for atmospheric attenuation (Equations 5.4 and 5.5),

\[\text{http://www.aips.nrao.edu/aipsmemo.html}\]
Figure 5.8: Example of SEFD values during a single night of the 2017 EHT observations (April 11, low band RCP). Values for 3C279 are marked with full circles, values for M87 are marked with empty diamonds. ALMA SEFDs have been multiplied by 10 in this plot. The SPT is observing 3C279 at an elevation of just 5.8°, resulting in an uncharacteristically high SEFD due to the large airmass.

2. DPFU: the degrees per flux density unit provides the conversion factor (K/Jy) from a temperature scale to a flux density scale, correcting for the aperture efficiency (Equation 5.6),

3. \( \eta_{el} \): the gain curve is a modeled elevation dependence of the telescope’s aperture efficiency (Equation 5.7), factored out of the DPFU to track gain variation as the telescope moves across the sky.

The EHT is a heterogeneous array with telescopes of various sensitivities (ranging nearly three orders of magnitude, see Figure 5.8), operation schemes, and designs. A clear understanding of each station’s metadata measurement and delivery is required for an accurate calibration of the measured visibilities. We determine the SEFDs of the individual stations and their uncertainties under idealized conditions, assuming adequate pointing and focus (see Sections 5.6.1, 5.6.1 and 5.6.1). Further losses and uncertainty in the SEFDs, particularly those induced by focus or pointing errors, are difficult to quantify using available metadata, but are qualitatively explained in
Section 5.6.1. A more quantitative assessment of station behavior can be done via derived residual station gains from self-calibration methods in imaging or model fitting [Paper IV; Paper VI].

Quantifying Station Performance

In order to determine the sensitivity of a single-dish station at a given time, measurements of the effective system temperature, the DPFU, and the gain curve are required. Here we provide details on how these parameters are measured for the EHT array.

The EHT operates in the millimeter-wave radio regime, where observations are very sensitive to atmospheric absorption and water vapor content. In contrast with centimeter-wave interferometers (e.g., VLBA/JVLA), millimeter-wave telescopes typically measure $T_{\text{sys}}$ via the ‘chopper’ (or hot-load) method: an ambient temperature load $T_{\text{hot}}$ with known blackbody properties is placed in front of the receiver, blocking everything but the receiver noise, and the resulting noise power is compared to the same measurement on cold sky. Assuming $T_{\text{hot}} \sim T_{\text{atm}}$ (the hot load is at a temperature comparable to the radiating atmosphere), this method automatically compensates for atmospheric absorption to first order, essentially transferring the incident flux density reference point to above the atmosphere [e.g., Penzias & Burrus, 1973; Ulich & Haas, 1976]:

$$T_{\text{sys}}^* \simeq e^\tau (T_{\text{rx}} + (1 - e^{-\tau}) T_{\text{atm}}),$$

(5.4)

where $T_{\text{rx}}$ is the receiver noise temperature and $\tau$ is the sky opacity in the line of sight. Details on the chopper techniques adopted for the EHT are provided in a technical memo\footnote{EHT Memo Series: https://eventhorizontelescope.org/for-astronomers/memos.} [Appendix A; Issaoun et al., 2017a].

Three stations in the EHT array have double-sideband (DSB) receivers in 2017 (SMA, JCMT and LMT), where both upper and lower sidebands on either side of the oscillator frequency are folded together in the recorded signal (e.g., Iguchi 2005, Paper II). Because only one 4 GHz sideband is correlated across the array, we correct $T_{\text{sys}}^*$ for the excess noise contribution from the uncorrelated sideband,

$$T_{\text{sys}}^* = T_{\text{sys, DSB}}^* (1 + r_{\text{sb}}),$$

(5.5)

where the sideband ratio $r_{\text{sb}}$ is the ratio of source signal power in the uncorrelated sideband to that in the correlated sideband. A sideband ratio of unity, for an ideal DSB system, is assumed for the SMA and LMT based on known receiver performance. A measured sideband ratio of 1.25 is used for the JCMT\footnote{https://www.eaobservatory.org/jcmt/instrumentation/heterodyne/rxa/}. The remaining stations use sideband-separating receiver systems and do not need this adjustment. The SPT, although sideband-separating, is believed to have suffered from a degree of incomplete sideband separation in 2017, giving it some amount of (uncharacterized) effective $r_{\text{sb}}$.

In addition to the noise characterization, the efficiency of the telescope must also be quantified. The DPFU relates flux density units incident onto the dish to equivalent degrees of thermal noise power through the following equation:

$$\text{DPFU} = \frac{\eta A_{\text{geom}}}{2k_B},$$

(5.6)
Figure 5.9: Example of a gain curve fit to single-dish normalized flux density measurements of calibrators at the SMT [Appendix B; Issaoun et al., 2017b].

where $k_B$ is the Boltzmann constant ($k_B = 1.38 \times 10^3$ Jy/K), $A_{\text{geom}}$ is the geometric area of the dish, and $\eta_A$ is the aperture efficiency of the telescope. For an idealized telescope with a uniform illumination (no blockage or surface errors), the full area would be available to collect the incoming signal and the aperture efficiency would be unity. Real radio telescopes intentionally taper their illumination to minimize spillover past the primary mirror, most have secondary mirror support legs that block part of the primary aperture, and generally the surface accuracy produces a non-negligible degradation in efficiency. To determine $\eta_A$, well-focused and well-pointed observations are made of calibrator sources of known brightness, usually planets [e.g., Kutner & Ulich, 1981; Mangum, 1993; Baars, 2007]. The planet brightness temperature models from the GILDAS\textsuperscript{14} software package were used for this calibration. For each single-dish EHT station, we determine a single DPFU value per polarization/band, except for JCMT, which has measurable temporal variations from solar heating during daytime observations. A more detailed overview of the methodology for $\eta_A$ is presented in Issaoun et al. [2017a] (Appendix A).

\textsuperscript{14}http://www.iram.fr/IRAMFR/GILDAS.
Table 5.2: Median EHT station sensitivities on primary targets during the 2017 campaign, assuming nominal pointing and focus.

<table>
<thead>
<tr>
<th>Station</th>
<th>Diameter in 2017 (m)</th>
<th>Sideband Ratio</th>
<th>Sideband-corrected Median $T_{\text{sys}}^*$ (K)</th>
<th>Aperture Efficiency $\eta_{A}$ (K/Jy)</th>
<th>DPFU (K/Jy)</th>
<th>Multiplicative Mitigation Factor</th>
<th>Median SEFD (Jy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APEX</td>
<td>12</td>
<td>—</td>
<td>118</td>
<td>0.61</td>
<td>0.025</td>
<td>1.020</td>
<td>4800</td>
</tr>
<tr>
<td>JCMT</td>
<td>15</td>
<td>1.25</td>
<td>345</td>
<td>0.52</td>
<td>0.033$^a$</td>
<td>—</td>
<td>10500</td>
</tr>
<tr>
<td>LMT</td>
<td>32.5</td>
<td>1.0</td>
<td>371</td>
<td>0.28</td>
<td>0.083</td>
<td>—</td>
<td>4500</td>
</tr>
<tr>
<td>PV</td>
<td>30</td>
<td>—</td>
<td>226</td>
<td>0.47</td>
<td>0.12</td>
<td>3.663</td>
<td>6900</td>
</tr>
<tr>
<td>SMT</td>
<td>10</td>
<td>—</td>
<td>291</td>
<td>0.60</td>
<td>0.017</td>
<td>—</td>
<td>17100</td>
</tr>
<tr>
<td>SPT</td>
<td>6$^b$</td>
<td>—</td>
<td>118</td>
<td>0.60</td>
<td>0.0061</td>
<td>—</td>
<td>19300</td>
</tr>
<tr>
<td>SMA6</td>
<td>14.7$^c$</td>
<td>1.0</td>
<td>285</td>
<td>0.75</td>
<td>0.046$^d$</td>
<td>1.138$^e$, 1.515$^e$</td>
<td>6400</td>
</tr>
<tr>
<td>ALMA37</td>
<td>73$^c$</td>
<td>—</td>
<td>76</td>
<td>0.68</td>
<td>1.03$^d$</td>
<td>—</td>
<td>74</td>
</tr>
</tbody>
</table>

Notes.
$^a$ Nighttime value for the DPFU. The daytime DPFU includes a Gaussian component dip as function of local Hawaii time.
$^b$ SPT has a 10 m dish diameter, with 6 m illuminated by receiver optics in 2017.
$^c$ The diameter for phased arrays reflects the sum total collecting area.
$^d$ DPFUs for phased arrays are determined for the full collecting areas.
$^e$ Applied when 6.25% and 18.75% of the SMA bandwidth was corrupted respectively.
We separately determine the elevation-dependent efficiency factor $\eta_{el}$ (or gain curve) due primarily to gravitational deformation of each parabolic dish. The characterization of the telescope’s geometric gain curve is particularly important for the EHT, which often observes science targets at extreme elevations in order to maximize $(u,v)$ coverage. The elevation-dependent gain curve is estimated by fitting a second-order polynomial to measurements of bright calibrator sources continuously tracked over a wide range of elevation (see Figure 5.9 and the technical memo\textsuperscript{12} by Issaoun et al. 2017b). In the EHT array, SMT, PV and APEX have characterized gain curves. The gain curve is parameterized as a second-order polynomial about the elevation at maximum efficiency:

$$\eta_{el} = 1 - B (el - el_{\text{max}})^2.$$  \hspace{1cm} (5.7)

The JCMT has no elevation dependence at 230 GHz as it is operating at the lower end of its frequency range. The LMT has an adaptive surface that is able to actively correct for surface deformation as a function of elevation. Through observations of planets, the LMT was determined to have a flat 1.3 mm gain between 25–80° to within 10% uncertainty. At the SPT, the elevation of extra-solar sources is constant, and therefore possible elevation-dependent efficiency losses remain uncharacterized.

We also mitigate a number of pathological issues uncovered in the 2017 data affecting the visibility amplitudes in a priori calibration. Additional loss of coherence in the signal chain at PV due to impurities in the local oscillator, an excess noise contribution at APEX due to the inclusion of a timing signal, and the partial SMA channel dropouts were identified during data processing. Correction factors for the visibility amplitudes on baselines to these sites were estimated, as explained in Appendix 5.A. These correction factors translate to a square multiplicative effect on the station-based SEFDs, as shown in Table 5.2. In the a priori calibration metadata, the multiplicative factors were folded into the DPFUs for PV and APEX and into the $T_{\text{sys}}^*$ measurements for SMA (due to its time-dependence). Representative median values for the aperture efficiency, DPFU, effective system temperature, and SEFD on EHT primary targets (M87 and Sgr A*) for each station participating in the EHT 2017 observations are shown in Table 5.2. A site-by-site overview of the derivation of a priori calibration quantities is given in a technical memo\textsuperscript{12} [Janssen et al., 2019a].

**Calibrating Visibility Amplitudes**

The $T_{\text{sys}}^*$, DPFU and elevation gain data for all stations are aggregated in ANTAB format text files. They are subsequently matched with observed visibilities for a given source using linear interpolation. Visibility amplitudes are calibrated in units of flux density by multiplying the normalized visibility amplitudes by the geometric mean of the derived SEFDs of the two stations across a baseline $i$–$j$:

$$|V_{ij}| = \sqrt{\text{SEFD}_i \times \text{SEFD}_j} |r_{ij}|,$$  \hspace{1cm} (5.8)

where $|V_{ij}|$ is then the calibrated visibility amplitude in Jy on that baseline, as in Equation 5.2.

Figure 5.10 shows the scan-averaged S/N on individual baselines, which is proportional to the phase-calibrated correlated signal, as a function of the projected baseline length (top panel),
Figure 5.10: Stages of visibility amplitude calibration illustrated with the April 11 HOPS data set on M87 (left) and 3C279 (right), as a function of projected baseline length. The two frequency bands are coherently scan-averaged separately and the final amplitudes are averaged incoherently across bands. Top: S/N of the correlated flux density component after phase calibration, both RCP and LCP. Middle: flux-density calibrated RCP and LCP values. Bottom: final, network calibrated Stokes I flux densities. Error bars denote ±1σ uncertainty from thermal noise.
and the equivalent correlated flux density after a priori calibration (center panel) for observations of M87 (left) and 3C279 (right) on April 11. The split in the S/N distributions is due to the difference in sensitivity between the co-located sites ALMA and APEX, leading to simultaneous baselines with two levels of sensitivity. The a priori calibration process puts all points on the same flux density scale (via Equation 5.8), and the resulting data variations can thus be attributed to source structure, no longer dominated by sensitivity differences between baselines.

**Single-Dish Error Budget**

The SEFD error budget, assuming nominal pointing and focus, is dominated by the measurement uncertainty for the DPFU (see Table 5.3). Depending on the source elevation, the uncertainty contribution for the elevation gain may also be non-trivial (particularly for the LMT) and adds in quadrature to the DPFU error to give the SEFD error budget. The gain curve error budget is obtained from the propagation of errors on the polynomial fit parameters in Equation 5.7, and is also itself elevation-dependent. We assume that the uncertainty in $T_{\text{sys}}^*$ is negligible as it is the variable measured closest to the individual VLBI scans and the accuracy of the chopper method is well studied [see Section 5.6.1, Kutner, 1978; Mangum, 2002]. The measurement uncertainties associated with pointing or focus errors are not folded into these error budget estimates as they are not easily quantifiable a priori.

For all single-dish stations, the DPFU uncertainty is estimated by the standard deviation in $\eta_A$ from a distribution of planet measurements added in quadrature to the uncertainty in the model brightness temperatures assumed for the planets. The scatter in planet measurements reflects changes in telescope performance with varying weather conditions, and thus it encompasses possible fluctuations in the mean value assumed during the observing window. An exception is the JCMT during daytime observing, where $\eta_A$ has a time-dependence parametrized by a fit of a Gaussian component dip as a function of local time, described in a technical memo [Issaoun et al., 2018]. The uncertainty in $\eta_A(t)$ is determined through the propagation of the errors on the fit parameters via least-squares fitting. Individual uncertainty contributions of the various components and the resulting percentage SEFD error budget for each EHT station during the 2017 April observations are listed in Table 5.3. Site-by-site derivations of flux density calibration uncertainties during the EHT 2017 campaign are given in Janssen et al. [2019a].

**Phased Array Calibration**

The phased arrays combine the total collecting area of all their dishes into one virtual telescope. This depends on precise phase-alignment of the signals, with an accuracy that is captured by the phasing efficiency $\eta_{\text{ph}}$ [Paper II, Appendix A],

$$\eta_{\text{ph}} = \frac{|\sum \gamma_i|^2}{(|\sum |\gamma_i|)^2}. \quad (5.9)$$

The phasing efficiency contributes to the aperture efficiency of the phased array, and reflects the ratio of source signal power\(^{15}\) observed by the phased array, versus that observed by a perfectly

\(^{15}\)It is common to see $\eta_{\text{ph}}^{1/2}$ defined as the phasing efficiency [e.g., Matthews et al., 2018], which scales with signal amplitude.
5.6 Flux Density Calibration

Table 5.3: Station-based SEFD percentage error budget during the 2017 campaign, assuming stable weather conditions and nominal pointing and focus. Subdominant effects from $T_{sys}$ measurements and sideband ratios are not shown.

<table>
<thead>
<tr>
<th>Station</th>
<th>DPFU</th>
<th>Gain Curve</th>
<th>$\eta_{ph}$</th>
<th>SEFD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Budget (%)</td>
<td>Budget (%)</td>
<td>Budget (%)</td>
<td>Budget (%)</td>
</tr>
<tr>
<td>APEX</td>
<td>11</td>
<td>0.3</td>
<td>—</td>
<td>11</td>
</tr>
<tr>
<td>JCMT</td>
<td>11–14$^a$</td>
<td>—</td>
<td>—</td>
<td>11–14</td>
</tr>
<tr>
<td>LMT</td>
<td>20</td>
<td>10</td>
<td>—</td>
<td>22$^b$</td>
</tr>
<tr>
<td>PV</td>
<td>10</td>
<td>1.5</td>
<td>—</td>
<td>10</td>
</tr>
<tr>
<td>SMT</td>
<td>7</td>
<td>1</td>
<td>—</td>
<td>7</td>
</tr>
<tr>
<td>SPT</td>
<td>15</td>
<td>—</td>
<td>5–15$^c$</td>
<td>15$^b$</td>
</tr>
<tr>
<td>SMA6</td>
<td>2</td>
<td>—</td>
<td>5–15</td>
<td>5–15</td>
</tr>
<tr>
<td>ALMA37</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>10$^d$</td>
</tr>
</tbody>
</table>

Notes. —

$^a$ The range in the budget at the JCMT is due to a larger uncertainty in the calibration during daytime observing, due to its aperture efficiency time-dependence.

$^b$ The error budget for SPT and LMT are lower limits due to uncharacterized losses, see Section 5.6.1.

$^c$ The range in the budget at the SMA is due to a larger uncertainty in the phasing for weaker sources.

$^d$ ALMA uncertainty is a lower limit from systematics caused by the assumed source flux density during QA2 calibration.

phased array. The complex gains $\gamma_i$ (as in Equation 5.2) are taken over all the dishes in the phased array, and have zero relative phase in the case of ideal phasing ($\eta_{ph} = 1$).

The phasing efficiency as defined above is valid when the signals being combined are optimally weighted by the effective collecting area of each antenna, $A_{i,eff} \sim 1/\text{SEFD}_i$. Then the SEFD of the phased array is

$$\text{SEFD}_{array} = \frac{1}{\eta_{ph}} \left( \sum \frac{1}{\text{SEFD}_i} \right)^{-1}. \quad (5.10)$$

Both SMA and ALMA use equal weights for the formation of the sum signal. Due to their homogeneity, Equations 5.9 and 5.10 are excellent approximations.

At the SMA, the phasing efficiency $\eta_{ph}$ is estimated from self-calibrated phases to a point-source model [Young et al., 2016]. Phases for each dish of the connected-element array are calculated online once per integration period, which varies in the range of 6–20s depending on the observing conditions, and the same phases are fed back as corrective phases for beamforming the phased array. The DPFU for the individual antennas that comprise the SMA are well characterized at 0.0077 K/Jy, with $\eta_A = 0.75$, and the 6m dishes have a flat gain curve at 230 GHz, which is near the lower end of their operating frequency range [Matsushita et al., 2006]. An SEFD for each antenna is calculated from DSB $T_{sys}^*$ measurements taken regularly at the time of observing. The overall SEFD for the SMA phased array is then estimated via Equation 5.10.

For ALMA, both amplitude and phase gain for each dish are solved during the offline QA2 processing of interferometric ALMA data, under an assumed point-source model with known total
Chapter 5: EHT Data Processing and Calibration

flux. The SEFDs of individual antennas are thus determined through amplitude self-calibration, automatically accounting for system noise and efficiency factors but sensitive to errors in the source model. Because ALMA data has the additional complication of linear-to-circular conversion, the phased-sum signal SEFD is determined via the full-Stokes Jones matrix of the phased array, as computed by PolConvert [Equation 15 of Martí-Vidal et al., 2016a]. By convention, QA2 sensitivity tables place all phasing related factors into the $T_{\text{sys}}$ component of Equation 5.3, allowing DPFU to assume a constant value corresponding to a single ALMA antenna. Further details are provided in Section 6.2.1 of Goddi et al. [2019].

During the EHT 2017 observations, $\eta_{ph}$ was above 0.8 for $\sim$80% (ALMA) and $\sim$90% (SMA) of the time. Poorer efficiency at both sites is associated with low elevation and increased atmospheric turbulence. At ALMA, phase corrections are calculated online by the telescope calibration system and applied to the array with a loop time of $\sim$18 seconds [Goddi et al., 2019]. At the SMA, integration times at the correlator can be as short as 6 seconds, but longer intervals are used if needed to build S/N. The corrective phases are passed through a stabilization filter before being applied, resulting in an effective loop time of $\sim$12–40s for the SMA. Phasing at both sites suffers when the atmospheric coherence timescale becomes short with respect to the loop time. To minimize the impact, both arrays are arranged in tight configurations during phased array operations.

The uncertainty on the $\eta_{ph}$ measurement at the SMA is estimated to be 5–15%, and depends primarily on the S/N of the gain solutions. The SMA (six 6m dishes$^5$) has considerably less collecting area than ALMA (37 12m dishes$^5$) to use for solving phase gains. For weaker sources, the uncertainty in estimating corrective phases at the SMA and in calculating the phasing efficiency can be considerable. The assumed flux of the point-source model used to self-calibrate ALMA during QA2 has a quoted 10% systematic uncertainty in Goddi et al. [2019]. The uncertainties from self-calibration and phasing are uncharacterized, therefore the uncertainty of 10% for the derived SEFD of the ALMA phased array is considered a lower limit. Errors from the use of a point-source model for M87 and 3C279 during gain calibration are expected to be small in comparison to these values. The individual uncertainties and error budget for the phased arrays are shown in Table 5.3.

Limitations of A Priori Calibration

Although the DPFU is typically represented as a single value measured under good performance conditions, a station’s efficiency is expected to vary with temperature, sunlight, and quality of pointing and focus. We have attempted to characterize specific time-dependent trends such as daytime dependence for the JCMT, but other factors are very difficult to decouple from the overall station behaviour and to associate with individual scans. Specific efficiency losses during scans, in particular due to lack of pointing/focus accuracy, are not included in the a priori amplitude calibration information for single-dish sites and remain in the underlying correlated visibilities. Therefore, the a priori error budget in Table 5.3 is only representative of global station performance and cannot be estimated for individual scans. In addition to a priori calibration, a list of problematic scans, where the station performance is known to be poor and the error budget is thus assumed to be undetermined, is passed on to analysis groups. These losses can be corrected
5.6 Flux Density Calibration

in imaging and model fitting via self-calibration methods and amplitude gain modeling [Paper IV; Paper VI].

The uncertainty in the chopper calibration is also difficult to quantify, as we do not know the true coupling of the hot load to the receiver (including spillover and reflection) and thus its effective temperature is uncertain [Kutner, 1978; Jewell, 2002]. One of the key assumptions of the chopper method is the equivalence (to first order) of the hot load, ambient and atmospheric temperatures, which allows for the correction of the atmospheric attenuation in the signal chain. Any deviation from this assumption in the $T_{sys}$ measurements may introduce systematic biases. This can be partly mitigated by frequent measurements and monitoring of the DPFU under stable weather conditions and nominal telescope performance, to offset any significant scaling from temperature assumptions. The majority of stations in the EHT use a two-load (hot and cold loads) chopper method, with temperature refinement from atmospheric modeling, to measure the receiver noise temperature, and have radiometers to monitor the atmospheric opacity, which typically reduces uncertainty in the chopper calibration down to the 1% level [Jewell, 2002; Mangum, 2002]. In contrast, the LMT and SPT used a single-load chopper method in 2017, leading to a larger error contribution estimated at the 5–10% level minimum [Jewell, 2002; Mangum, 2002]; with an error that grows rapidly at high line-of-sight opacity.

Limitations in accuracy of the a priori calibration may also come from the cadence of DPFU and $T_{sys}$ measurements, typically performed between scheduled VLBI scans or outside VLBI observing altogether. The changing dish performance during the VLBI observations and intra-scan atmospheric variations are not typically captured by these measurements, although frequent pointing and focus calibration is done during the observations to keep an optimal performance. Furthermore, the time cadence varies across participating stations due to different chopper calibration setups, pointing and focus needs, and allocated time for the EHT observing campaign. It is therefore not atypical for self-calibration corrections in downstream analysis to slightly deviate from the attributed amplitude error budget. To maximize mutual coverage, many stations are pushed past their nominal operating conditions during EHT observations, such as the LMT or the JCMT in the early evening local time due to surface heating and instability, and the SPT at extremely low elevation and high winds. For those stations and conditions, we expect residual gains to deviate significantly from the a priori amplitude error budget. A more detailed discussion of a priori calibration uncertainties and limitations is given in Issaoun et al. [2017a].

5.6.2 Network Calibration

Network calibration is a framework to estimate visibility amplitude corrections at some sites by utilizing array redundancy and supplemental measurements of the total flux density of a source [Fish et al., 2011; Johnson et al., 2015; Blackburn et al., 2019]. It allows for absolute amplitude calibration of intra-site baselines and tightens consistency between simultaneous baselines to colocated sites when both sites are observing (see bottom panels of Figure 5.10). It makes fewer assumptions than other techniques such as self-calibration and does not assume a specific compact source model.

Network calibration makes two related assumptions. The first is that redundant baselines in the EHT array (e.g., ALMA–SMA and APEX–JCMT) share the same model visibility. The
second is that co-located sites provide a zero-baseline interferometer (e.g., ALMA–APEX), with a corresponding visibility that is a positive real number equal to the total flux density $V_0$. We express the measured visibility $V_{ij}$ on a baseline between sites $i$ and $j$ as

$$V_{ij} = g_i g_j^* V_{ij}, \quad (5.11)$$

where $V_{ij}$ is the true visibility on that baseline, and $g_i$ and $g_j$ are the station-based residual gains assuming no thermal noise (the latter introduces uncertainty in the estimated gains).

Given two co-located sites $i$ and $j$, we can solve for the amplitudes of their gains using a third remote site, using the assumptions above, $V_{ik} = V_{jk}$ and $V_{ij} = V_0$. In the absence of thermal noise,

$$|g_i| = \sqrt{\frac{V_{ij}}{V_0} \times \frac{V_{ik}}{V_{jk}}} \quad \text{and} \quad |g_j| = \sqrt{\frac{V_{ij}}{V_0} \times \frac{V_{jk}}{V_{ik}}}. \quad (5.12)$$

Note that network calibration only provides gain estimates for those sites with a co-located partner.

In practice, thermal noise affects the accuracy of gains estimated using Equation 5.12. To optimize network calibration, we use all sets of baselines between co-located sites and distant sites and solve for the set of unknown model visibilities $V_{ij}$ and station gains $g_j$ by minimizing an associated $\chi^2$. Specifically, for each solution interval, we minimize

$$\chi^2 = \sum_{i<j} \frac{|g_i g_j^* V_{ij} - V_{ij}|^2}{\sigma_{ij}^2}, \quad (5.13)$$

where $\sigma_{ij}$ is the thermal uncertainty on $V_{ij}$. We implemented network calibration via this minimization procedure within the eht-imaging library [Chael et al., 2016, 2018b].

For the EHT April 2017 observations, network calibration is performed on frequency-averaged visibility UVFITS data coherently time-averaged over ten second solution intervals. Both parallel hand visibility components (further referred to as RCP/LCP or RR/LL) are network-calibrated with shared gain coefficients, using the total intensity measured by the ALMA array as $V_0$ [Goddi et al., 2019]. The assumed flux density values per band on each observing day are reported in Table 5.4 for both M87 and 3C279. For each source, a constant flux density is adopted per day, as both sources vary by $<5\%$ within an observation, well within the 10\% flux density calibration error budget of ALMA measurements.

Network calibration enables absolute amplitude calibration of sites with a co-located partner (ALMA & APEX, SMA & JCMT) when both sites are operating, to the limit of thermal noise to the strongest remote stations. The remaining isolated sites (SMT, LMT, SPT, PV) are unaffected by network calibration.

Following all calibration steps, Stokes $I$ total intensity components correspond to

$$V_{ij,I} = \frac{1}{2} (V_{ij,RR} + V_{ij,LL}). \quad (5.14)$$

For JCMT, which is a single polarization station, we use the available RCP or LCP component as a proxy for the Stokes $I$ value. This corresponds to assuming zero contribution from Stokes $V$ circular polarization.
Table 5.4: Total flux density estimates used for network calibration.

<table>
<thead>
<tr>
<th>Source</th>
<th>Band</th>
<th>April 5 (Jy)</th>
<th>April 6 (Jy)</th>
<th>April 10 (Jy)</th>
<th>April 11 (Jy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M87</td>
<td>low</td>
<td>1.13</td>
<td>1.14</td>
<td>1.17</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>1.12</td>
<td>1.10</td>
<td>1.15</td>
<td>1.20</td>
</tr>
<tr>
<td>3C279</td>
<td>low</td>
<td>8.61</td>
<td>8.57</td>
<td>7.99</td>
<td>8.01</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>8.56</td>
<td>8.55</td>
<td>7.97</td>
<td>7.98</td>
</tr>
</tbody>
</table>

Note. – The flux density values used for network calibration in SR1 come from the initial ALMA QA2 data release (October 2017), with a quoted uncertainty of 10%. Updated values are reported in Appendix B of Goddi et al. [2019] and are approximately 10% higher than shown here.

Most assumptions in the network calibration procedure are valid for all targets observed by the EHT. However, the assumption that co-located sites act as a true zero-baseline interferometer may not hold for sources with extended structure, such as M87. The distance between the SMA and the JCMT is 160 m, giving a resolution on that baseline of \( \sim 1.6'' \). The distance between ALMA (phase center) and APEX is 2.6 km, giving a resolution on that baseline of \( \sim 0.1'' \). For very compact sources, such as the quasar 3C279, these two baselines both see point-like sources. For sources with extended structure, such as M87 and its large-scale jet, these two baselines will see slightly different structure. For example HST–1, a bright feature in the jet of M87 at just 0.8'' from the radio core [Chang et al., 2010], produces a different response on both intra-site baselines. However, HST–1 has \( \leq 1\% \) of the total core flux density of M87 as measured by ALMA (Table 5.4), so its effect on the network calibration gain solutions for ALMA and APEX is insignificant in comparison to the 10% uncertainty on the ALMA total flux density estimates.

5.7 Final Data Products

5.7.1 Data Release Specification

The SR1 data on M87 and 3C279 represent a subset of a more comprehensive engineering data production (ER5) for the EHT 2017 observations, after extensive internal validation and review. ER5 data are themselves derived from a fifth revision (Rev5) correlation data product.

The sequence of correlation and engineering releases represents a year-long effort of identifying and mitigating data issues, and developing new software and procedures; first on secondary targets for ER1–ER3 and then including EHT primary science targets for ER4–ER5. Each internal engineering data release was subject to an independent review by a panel of experts not involved in the data preparation, before being made available for downstream analysis, including imaging and model fitting. The HOPS data set was present in all engineering releases, receiving the most extensive review and internal validation. AIPS data were included in ER1 for an initial comparison to HOPS on EHT 2017 secondary targets, and in ER5 for comparisons with both HOPS and the newly added CASA data set.

The final data products at the end of the calibration and reduction pipelines provide a uniform and reliable data set for scientific analysis that has been reduced and simplified by the removal
Chapter 5: EHT Data Processing and Calibration

Table 5.5: Data products available in SR1.

<table>
<thead>
<tr>
<th>Stage</th>
<th>( \Delta t ) (s)</th>
<th>( \Delta \nu ) (MHz)</th>
<th>Low Band (GB)</th>
<th>High Band (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corr. Data (Rev5)</td>
<td>0.4</td>
<td>0.5</td>
<td>665</td>
<td>713</td>
</tr>
<tr>
<td>Phase Cal. (SR1)</td>
<td>0.4</td>
<td>58.0</td>
<td>7.9</td>
<td>8.0</td>
</tr>
<tr>
<td>A Priori Cal. (SR1)</td>
<td>0.4</td>
<td>58.0</td>
<td>7.9</td>
<td>8.0</td>
</tr>
<tr>
<td>Network Cal. (SR1)</td>
<td>10.0</td>
<td>1875.0</td>
<td>0.117</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Note – Integration time \( \Delta t \) and frequency averaging windows \( \Delta \nu \) are given, as well as total data volumes for low and high band subsets, which have slightly different coverage.

of bad data (failed observations), and after compensating for non-astrophysical systematics. The data reduction process is automated and makes only minimal assumptions about the source: (1) that the target is mostly compact, and (2) that it has known a priori large-scale structure and total flux density (e.g., from ALMA observations). The calibration of systematics is therefore limited by an inability to jointly fit source parameters along with gains, but this pathway avoids introducing any strong model assumptions during the data preparation.

In addition to the raw correlator output, three levels of successive data reduction are provided, representing the assumptions made during calibration. The first level (1) includes only the phase calibration provided during fringe fitting, after which data can be averaged. At this stage, the data represent correlation coefficients and are the most fundamental data product for the formation of closure phases and closure amplitudes. This is followed by (2) data that has been brought to a physical amplitude scale (Jy) through a priori flux density calibration, and then (3) network amplitude calibrated using a priori assumptions about large-scale source structure and total flux density. The time-frequency resolutions of the various data products are presented in Table 5.5, and generally exceed what is needed to capture source structure. This resolution is chosen to allow for a manageable data volume while still providing flexibility for downstream time-frequency averaging as well as the fitting of any residual systematics through additional model-dependent techniques such as self-calibration.

The SR1 data release includes products of all three fringe fitting pipelines. The HOPS pipeline data product is designated as the primary scientific EHT data set, given the degree of vetting it has received during an iterative process of five engineering releases and a current performance advantage at low S/N. The CASA and AIPS data sets are used for validation, including direct data cross-comparisons as well as validation of downstream analysis results. Each data product is provided in UVFITS format. The choice of format was motivated by the need for common output across all pipelines, and easy loading, inspection and imaging in all software used in the downstream analysis efforts and via readily available Python modules. A suite of metadata accompany the release, such as the ANTAB tables used for a priori calibration, documentation and validation tests for each processing and calibration stage, assessment of derived calibration solutions, and suggested flagging information from investigations of station performance.

The first science release only provides calibrated Stokes \( I \) (total intensity) products for M87 and 3C 279. A summary of the data set content and S/N statistics is shown in Table 5.6, and
a cumulative histogram of the Stokes $I$ component S/N in the fully averaged data set is shown in Figure 5.11. A median reported thermal uncertainty is about 7 mJy on non-ALMA baselines and remarkably only about 0.7 mJy on baselines to ALMA for Stokes $I$ single-band scan-averaged visibilities. In this first science release, the issue of polarimetric leakage calibration and correction is not addressed. Leakage has a relatively small influence on the total intensity and it is sufficient to parametrize the effects of leakage as a systematic source of non-closing errors (see Section 5.8). Future EHT results concerning polarimetry and other Stokes components will necessarily involve leakage calibration.

5.7.2 Closure Quantities

While the data release consists of reduced complex visibilities, derivative closure data products are particularly important for downstream data analysis, as well as for the description of data uncertainties. Unlike complex visibilities, closure quantities are robust against station-based gain errors. They are however susceptible to systematic non-closing errors, discussed in Section 5.8. For the needs of this paper, we only provide brief definitions and description of conventions.

We define a closure phase formed from baseline visibilities on a closed triangle $ijk$ as

$$
\psi_{C,ijk} = \text{Arg}(V_{ij}V_{jk}V_{ki}),
$$

with a corresponding uncertainty

$$
\sigma_{\psi_{C,ijk}} \approx \sqrt{S_{ij}^{-2} + S_{jk}^{-2} + S_{ki}^{-2}},
$$

where $S_{ij}$ is the estimated S/N, associated with the $V_{ij}$ visibility, that is

$$
S_{ij} = \frac{|V_{ij}|}{\sigma_{ij}}.
$$

Formation of closure phase cancels the station-based gain factors that appear in Equation 5.2. In the case of visibility amplitudes, the gain factors can be similarly canceled by the formation of the log closure amplitude, defined as

$$
\ln A_{C,ijk\ell} = \ln \frac{A_{ij}A_{k\ell}}{A_{ik}A_{j\ell}},
$$

for a quadrangle $ijk\ell$, where ‘ln’ is a natural logarithm and $A_{ij}$ represents debiased amplitude,

$$
A_{ij} = \sqrt{|V_{ij}|^2 - \sigma_{ij}^2}.
$$

The associated uncertainty of log closure amplitude is

$$
\sigma_{\ln A_{C,ijk\ell}} \approx \sqrt{S_{ij}^{-2} + S_{k\ell}^{-2} + S_{ik}^{-2} + S_{j\ell}^{-2}}.
$$

Uncertainties reported in Equations 5.16 and 5.20 are calculated based on propagation of thermal visibility errors and are strictly correct in a high S/N limit, where distributions of both types of closure quantities are well approximated with a normal distribution. The number of closure quantities that can be derived from SR1 visibilities is given in Table 5.6. The numbers describe a fully averaged (i.e., scan and 4 GHz band-averaged) dataset. We give the number of all closure
Table 5.6: Content of the SR1 data set.

<table>
<thead>
<tr>
<th></th>
<th>HOPS</th>
<th>CASA</th>
<th>AIPS</th>
<th>Shared</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>M87 scans</td>
<td>72</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>72</td>
</tr>
<tr>
<td>detections</td>
<td>771</td>
<td>753</td>
<td>706</td>
<td>702</td>
<td>898</td>
</tr>
<tr>
<td>median S/N</td>
<td>31.4</td>
<td>27.3</td>
<td>25.9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(shared set)</td>
<td>36.6</td>
<td>31.8</td>
<td>26.4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>all closure phases</td>
<td>912</td>
<td>889</td>
<td>790</td>
<td>784</td>
<td>1130</td>
</tr>
<tr>
<td>(non-redundant)</td>
<td>482</td>
<td>470</td>
<td>432</td>
<td>429</td>
<td>579</td>
</tr>
<tr>
<td>all closure amps</td>
<td>1938</td>
<td>1890</td>
<td>1569</td>
<td>1557</td>
<td>2520</td>
</tr>
<tr>
<td>(non-redundant)</td>
<td>410</td>
<td>399</td>
<td>361</td>
<td>358</td>
<td>507</td>
</tr>
</tbody>
</table>

3C279 scans

|              | 71    | 71    | 68    | 68    | 71   |
| detections   | 954   | 937   | 972   | 913   | 1246 |
| median S/N   | 250   | 219   | 187   | –     | –    |
| (shared set) | 259   | 230   | 213   | –     | –    |
| all closure phases | 1313 | 1285 | 1370 | 1258 | 1918 |
| (non-redundant) | 631 | 618 | 646 | 607 | 864  |
| all closure amps | 3342 | 3276 | 3591 | 3207 | 5361 |
| (non-redundant) | 560  | 547   | 578   | 536   | 793  |

Note – Data products in the fully averaged SR1 data set. The shared data set is composed of only those detections that are reported by all three pipelines. The max data set is a theoretical maximum calculated assuming perfect realization of the observation schedules. The full set of all closure quantities is shown, which is used to estimate systematics in Section 5.8; as well as the non-redundant set, which reflects the actual number of unique phase and amplitude degrees-of-freedom measured by the (uncalibrated) array.

quantities, corresponding to the full (or maximal) set formed from all possible loops over three or four stations in every scan. The full set has a balanced representation of baselines, and is used to estimate systematic errors in Section 5.8.4. Elements of a maximal set are, however, not independent (the set is highly redundant). We also provide the number of closure products in the non-redundant (or minimal) set. This is a reduced subset that captures all the available information in the closure quantities. Selection of a particular non-redundant dataset is not unique and in general non-trivial [Blackburn et al., 2020].

When intra-site baselines are present in the array, a special set of trivial closure quantities can be formed. Such closure phases and log closure amplitudes are zero by construction, within statistical uncertainties. While they do not carry any direct information about the source compact structure, they are useful for network calibration (Section 5.6.2) and the characterization of uncertainties, presented in Section 5.8.

5.7.3 Data Features

Certain properties of the reduced data can be directly observed in the behavior of visibilities and closure quantities. The data indicate remarkable persistent features in the structure of the M87 compact emission, as well as source structural variability on a timescale of days. In this section
5.7 Final Data Products

Figure 5.11: Cumulative histogram of Stokes I S/N in the HOPS data set for all observations of M87 and 3C279, using fully averaged data. Solid curves represent baselines to ALMA, while the dashed curves show all other baselines.

Figure 5.12: Aggregate \((u, v)\) coverage for M87 (left) and 3C279 (right) for the April 2017 observations, comparable for all three pipelines. Co-located sites (ALMA/APEX and SMA/JCMT) result in redundant baselines. The dashed circles show baseline lengths corresponding to fringe spacings of 25 and 50 \(\mu\)as.
we give a rudimentary interpretation of these features. The implications of these basic features for the imaging, modeling and scientific interpretation of the source structure are explored in companion papers [Paper I; Paper IV; Paper V; Paper VI].

Figure 5.12 shows the aggregate baseline coverage for EHT 2017 observations of M87 and 3C279 via the HOPS pipeline. The coverage and data properties via the other two pipelines are comparable. Our shortest baselines are between co-located sites (SMA–JCMT and ALMA–APEX). These baselines are sensitive to arcsecond-scale structure, while our longest baselines are sensitive to microarcsecond-scale structure. For M87, the highest resolution (fringe spacing of 25 $\mu$as) is achieved in the east-west direction on baselines joining the Hawaii stations to PV, while for 3C279 the highest resolution (fringe spacing of 24 $\mu$as) is achieved in the north-south direction, on PV and SMT baselines to the SPT.

The 2017 observations led to detections on all baselines for M87. A longer averaging time (up to scan duration) is enabled by the atmospheric phase corrections performed by all three pipelines. Figure 5.10 (top left panel) shows the S/N as a function of projected baseline length for M87 on April 11, for fully averaged data. A similar distribution is also shown for 3C279 in Figure 5.10 (top right panel), with around an order of magnitude difference due to the higher total flux density of 3C279 compared to M87 (Table 5.4).

The correlated flux density for M87 on April 11 after amplitude and network calibration is shown in Figure 5.10 (bottom left panel). There is a pronounced secondary peak in the visibility amplitudes with two minima on either side, interpreted as visibility nulls. The first of these nulls occurs at $\sim$3.4 $G\lambda$. It is steep on the east-west oriented LMT and SMT baselines to the Hawaii stations, and shallower on the north-south oriented ALMA and APEX baselines to LMT at the same baseline length. The second null in amplitude is observed at $\sim$8.3 $G\lambda$, on the east-west oriented PV baselines to the Hawaii stations. The correlated flux density for 3C279 on April 11 after amplitude and network calibration is also shown in Figure 5.10 (bottom right panel). The trend in the visibility amplitudes is clearly different from the trend seen in M87. 3C279 appears to have more complex structure on long baselines, and the structure varies with baseline position angle.

Persistent Structural Features

Figure 5.13 shows the correlated flux density after amplitude and network calibration as a function of baseline length for all four days of observations of M87 via the HOPS pipeline. The network calibrated amplitudes show broad consistency over different days, and are consistent between pipelines (Section 5.8.5). The majority of notable low-amplitude outliers across days are due to reduced efficiency of the JCMT or the LMT on a select number of scans (caused by, e.g., telescope pointing issues or surface instability). Although the amplitudes of these data points are low, closure information remains stable and is unaffected by station gain. This is shown by comparing the erratic amplitudes on the LMT–SMT baseline in Figure 5.13 (cluster of points at about 1 $G\lambda$) with the smooth trends in closure phase for the ALMA–LMT–SMT triangle (Figure 5.14, top left) and in closure amplitude for the ALMA–LMT–APEX–SMT quadrangle (Figure 5.14, top right).

The secondary peak in amplitude and the location of the two nulls are persistent for all four days. These signatures in the visibility amplitudes suggest that the source is not changing...
Figure 5.13: Correlated flux density of M87 as a function of projected baseline length for all four days of observations, from HOPS data that has been fully averaged. Outliers are due to reduced performance of the LMT or the JCMT. Error bars denote ±1σ uncertainty from thermal noise.

Figure 5.14: A selection of M87 closure phases (left and central columns) and log closure amplitudes (right column) as a function of Greenwich Mean Sidereal Time (GMST) for all four observed nights from the HOPS data set. Plotted uncertainties denote ±1σ ranges from thermal noise in the fully-averaged data.
dramatically over several days, is compact with a characteristic spatial scale of $\lesssim 50\,\mu\text{as}$, and exhibits similar structure over a range of baseline position angle. Long baselines with various orientations lie in a stable trend along the second peak, and a minimum in amplitude at $3.4 \, \text{G}\lambda$ is seen on both the east-west and north-south oriented baselines.

While the overall trend may indicate a compact and nearly circularly symmetric structure stable in time, a more detailed inspection of the data set suggests the presence of a slight anisotropy, also made evident by multiple measurements of non-zero closure phase. This can be seen comparing the ALMA/APEX–LMT and SMA/JCMT–LMT amplitudes in Figure 5.10 (bottom left). Both baselines probe a $(u,v)$ distance of about $3.4 \, \text{G}\lambda$, but they have a very different, nearly perpendicular orientation (Figure 5.12). Flux density measured on the north-south oriented ALMA–LMT baseline is a few times larger than that for the east-west oriented SMA–LMT. These properties translate to striking source features in imaging and model fitting, presented in Paper IV and Paper VI respectively.

**Time Variability**

M87 was observed on the two consecutive nights of April 5/6 and again four nights later for the two consecutive nights of April 10/11. We observe clear indications of modest source evolution between the two pairs of nights, and broad consistency within each pair. The evolution can be seen particularly well in the behavior of robust closure quantities.

Across the full set of closure quantities, some closure phases formed by wide and open triangles (e.g., ALMA–LMT–SMA, Figure 5.14 bottom left) show different closure phase trends between the first pair of days and the second pair. Additionally, the east-west oriented LMT–SMA–SMT triangle shows different closure phase trends between the two pairs of days (Figure 5.14 bottom center), but the equivalent triangle in the opposite orientation, LMT–PV–SMT, shows no such trend (Figure 5.14 top center).

Strong night-to-night variability of closure phases is associated with baselines probing $(u,v)$ components close to the first visibility amplitude null, where visibility phases are particularly sensitive to small structural changes. The LMT–Hawaii baselines are particularly affected. Rapid swings of closure phase, as large as $200^\circ$ in 2 hours, are found for the LMT–SMA–SMT triangle, but exclusively for the latter pair of nights on April 10/11. Triangles that do not probe the $3.4 \, \text{G}\lambda$ null location indicate less variability, e.g., ALMA–LMT–SMT or LMT–PV–SMT. Despite larger uncertainties, similar trends are seen in log closure amplitudes (right column of Figure 5.14). In particular, significant differences between the two pairs of nights can be seen on the ALMA–LMT–APEX–SMA quadrangle, while the ALMA–LMT–APEX–SMT quadrangle gives more consistent values.

### 5.8 Data Validation and Systematics

In this section, we summarize data set validation tests, performed using diagnostic tools developed in the eat\textsuperscript{8} library framework and focusing on the properties of the final network-calibrated data products. The section is structured as follows. In Subsection 5.8.1, we discuss internal consistency tests performed during the fringe fitting stage. In Subsection 5.8.2, the accuracy
of reported thermal uncertainties is tested. In Subsection 5.8.3 we investigate the robustness of data products against decoherence with increased coherent averaging time. Subsection 5.8.4 presents internal consistency tests in each pipeline and provides estimates for the magnitude of non-closing systematic errors, which become important considerations in the error budget for high S/N measurements. Finally, in Subsection 5.8.5, direct comparisons between the three pipelines are given. A more comprehensive discussion of these automated data validation procedures is given in a technical memo\textsuperscript{12} [Wielgus et al., 2019].

5.8.1 Fringe Validation

During fringe detection, a number of basic tests are performed on the data that check for data integrity, false fringes, and the overall self-consistency of the detected fringe solutions and measured correlation coefficients. These fringe validation tests reflect the internal validation of each pipeline, as opposed to the overall statistical validation and cross-comparisons presented in following subsections. In addition to identifying issues with the fringe fitting pipelines themselves, consistent review of data products throughout engineering data production played an important role in characterizing upstream issues with the data and their correlation.

Figures 5.15 and 5.16 show two fringe solution consistency tests that are run as part of an automated test suite at each stage of the HOPS pipeline (Section 5.5.1, with details in Blackburn et al. 2019). In Figure 5.16, as well as in subsequent plots of distributions, the number of $3\sigma$ outliers and size of the tested sample for each source are provided. The dashed black curve indicates a standard normal distribution with zero mean and unity variance.

The HOPS pipeline baseline-based fringe solutions (prior to the global enforcement of fringe closure) show smooth evolution across each observing night and consistency across four polarization products, which are independently fit. Delay calibration assumes a constant RCP versus LCP delay offset per night at each station, which is verified by the stability of RR−LL delays to within thermal measurement error. Independently measured delay-rates between polarizations are also consistent to within thermal error. The lack of large-deviation outliers in these fringe solution consistency tests is a strong indication that there are no false fringes or corrupted measurements above the detection threshold.

5.8.2 Thermal Error Consistency

Thermal error plays an essential role in the VLBI uncertainties, both for the visibilities as well as for the derivative closure quantities, for which uncertainties are simply propagated from the visibility errors (Section 5.7.2). An accurate accounting of thermal noise is essential for deriving faithful model fitting uncertainties, and for correct noise debiasing in the case of incoherently averaged amplitudes [Rogers et al., 1995]. Fundamentally, thermal uncertainty $\sigma_{th}$ in the real and imaginary components of the dimensionless complex correlation coefficient $r_{ij}$ (Equation 5.1) can be estimated from first principles. Under the assumption of a stationary white noise process at each antenna,

$$\sigma_{th} = \frac{1}{\eta Q \sqrt{2 \Delta t \Delta \nu}},$$

(Equation 5.21)
Figure 5.15: Measured residual relative delays for selected M87 baselines on April 11, reported by the HOPS pipeline (Section 5.5.1) prior to explicit fringe closure. The top panel shows smooth delay trends over the night for both parallel hands, LL (dots) and RR (crosses). The bottom panel shows the sum of the delays on this closed triangle, which is consistent with the expected value of zero to within statistical errors. After fringe closure, RR and LL are set to the same delay, and closure delay is zero by construction.
where $\Delta t$ is the integration time, $\Delta \nu$ is the averaged bandwidth, and $\eta_Q$ is the factor that accounts for quantization efficiency. The thermal uncertainties reported by each pipeline depend on self-consistent tracking of scale factors through data conversion and calibration, as well as accounting for the data weights and bandpass response over the averaging windows in Equation 5.21.

The UVFITS file format formally associates a weight $w$ for each visibility measurement, with associated reported uncertainty $\sigma_{\text{rep}} \equiv 1/\sqrt{w}$. In the ideal case, $\sigma_{\text{rep}}$ properly represents thermal uncertainties, $\sigma_{\text{rep}} = \sigma_{\text{th}}$. For the HOPS and CASA pipelines, the thermal uncertainty is determined from first principles. However, the weights for the AIPS pipeline require a large scaling factor to be applied for their final output to ensure that $\sigma_{\text{rep}} = \sigma_{\text{th}}$. We derive this correction factor using the scatter from differences in adjacent high-S/N closure phases. For CASA, the direct interpretation of reported weights as $1/\sigma_{\text{th}}^2$ also leads to a small bias, resulting in underestimation of $\sigma_{\text{th}}$ by approximately 5%, as estimated by the closure phase differencing technique.

We test the scan-by-scan accuracy of $\sigma_{\text{rep}}$ via a comparison with an empirical estimator $\sigma_{\text{emp}}$, fitting the moments of visibility amplitudes distribution. We estimate $\sigma_{\text{emp}}$ for each scan, baseline, band, and polarization combination, by using moment matching of the visibility amplitude distribution over the scan duration [Wielgus et al., 2019]. Each ensemble is composed of, on average, 900 individual visibility amplitude measurements. Figure 5.17 shows distributions of $(\sigma_{\text{rep}} - \sigma_{\text{emp}})/\sigma_{\text{rep}}$ for all three SR1 processing pipelines, using the 5399 ensembles shared by pipelines. The median of each distribution (med) is given in the legend of Figure 5.17, and shows ensemble values roughly consistent with the alternative closure phase differencing test.

---

16 see AIPS Memo 103; [http://www.aips.nrao.edu/aipsmemo.html](http://www.aips.nrao.edu/aipsmemo.html)
5.8.3 Temporal Coherence After Calibration

All three data pipelines correct for changing visibility phase over scans, both in the correction for a linear drift via the delay-rate and in corrections for stochastic, station-dependent wander from atmospheric contributions (see Section 5.5). Although these corrections do not provide absolutely calibrated visibility phase, they eliminate differential wander on short timescales, allowing the visibilities to be coherently averaged for longer intervals than the atmospheric coherence time. An imperfect phase correction will lead to decoherence in the averages, which, in severe cases, may introduce non-closing amplitude errors.

To evaluate the performance of the phase correction algorithms, we compute two quantities for each scan: the amplitude $A_{\text{scan}}$ resulting from coherent averaging visibilities over the full scan (3–7 minutes) and subsequent debiasing (Equation 5.19), and the amplitude $A_{2s}$ obtained from 2 seconds coherently averaged visibility segments that were subsequently incoherently averaged.
over the full scan [Rogers et al., 1995; Johnson et al., 2015]. The ratio $A_{\text{scan}}/A_{2\text{s}}$ then quantifies the loss in amplitude from uncorrected phase fluctuations within scans.
### Table 5.7: Systematic errors in SR1 data set.

<table>
<thead>
<tr>
<th>Test</th>
<th>M87 HOPS</th>
<th>M87 CASA</th>
<th>M87 AIPS</th>
<th>3C279 HOPS</th>
<th>3C279 CASA</th>
<th>3C279 AIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR-LL closure phases (deg)</td>
<td>&lt;1.0(0.2)</td>
<td>&lt;1.0(0.2)</td>
<td>&lt;1.0(0.2)</td>
<td>1.9(1.1)</td>
<td>1.9(1.1)</td>
<td>2.1(1.2)</td>
</tr>
<tr>
<td>RR-LL log closure amplitudes (%)</td>
<td>&lt;2.0(0.2)</td>
<td>&lt;3.0(0.3)</td>
<td>&lt;2.0(0.2)</td>
<td>3.1(1.0)</td>
<td>3.6(1.2)</td>
<td>3.3(1.0)</td>
</tr>
<tr>
<td>Stokes $I$ closure phase low/high (deg)</td>
<td>1.4(0.4)</td>
<td>2.5(0.6)</td>
<td>2.6(0.6)</td>
<td>2.2(1.5)</td>
<td>2.3(1.5)</td>
<td>2.0(1.3)</td>
</tr>
<tr>
<td>Stokes $I$ log closure amplitude low/high (%)</td>
<td>5.6(0.8)</td>
<td>-</td>
<td>&lt;10.0(1.3)</td>
<td>4.5(1.8)</td>
<td>5.4(2.3)</td>
<td>4.8(1.8)</td>
</tr>
<tr>
<td>Stokes $I$ trivial closure phases (deg)</td>
<td>3.7(1.1)</td>
<td>2.6(0.8)</td>
<td>3.2(1.0)</td>
<td>1.2(1.9)</td>
<td>1.0(1.5)</td>
<td>1.0(1.4)</td>
</tr>
<tr>
<td>Stokes $I$ trivial log closure amplitudes (%)</td>
<td>3.6(0.4)</td>
<td>5.6(0.7)</td>
<td>7.7(0.9)</td>
<td>3.8(2.0)</td>
<td>3.8(1.9)</td>
<td>3.3(1.6)</td>
</tr>
</tbody>
</table>

Note – Characteristic magnitudes of systematic errors, estimated using the subset of data shared by all three pipelines. Scan-averaged single-band data. Numbers in parentheses represent characteristic systematic errors in units of thermal noise.
Figure 5.18 shows cumulative histograms of $A_{\text{scan}}/A_{2s}$ for a common subset of 4688 ensembles (subsets of unique scan, baseline, band, and polarization) shared between pipelines, with an $S/N > 7$ threshold. While small errors in the estimated thermal noise have little effect on the $S/N$ of coherent averages, they can significantly affect the outcome of incoherent averaging. Thus, only for this particular test, we applied a fixed correction factor of 1.05 to CASA thermal noise estimates $\sigma_{\text{rep}}$ before incoherent averaging, to account for the small bias in this pipeline discussed in Section 5.8.2. For all three pipelines, the coherence of the phase corrected data is significantly better than that of data with no atmospheric phase correction (the gray curve in Figure 5.18; see also Figure 2 of Paper II), with over 90% of the calibrated data experiencing an amplitude loss of under 10%. These results demonstrate that coherent averaging over scans is admissible for the SR1 data set, particularly in case of the HOPS data products.

5.8.4 Intra-Pipeline Validation

In this subsection we perform internal data consistency tests for each pipeline, in order to estimate the magnitude of systematic non-closing errors, e.g., related to the uncalibrated polarimetric leakage. For that purpose, we inspect closure phases and log closure amplitudes derived from the SR1 data set and evaluate consistency between (1) RR and LL components, (2) low and high frequency bands and (3) trivial closure quantities. For each test, we derive a magnitude of residual errors, in excess to the reported thermal uncertainties. These values are then used to characterize the magnitude of non-closing errors in the data set, utilized in the downstream analysis.

Quantifying Residual Errors

We evaluate the characteristic magnitude of systematic errors in the SR1 data set based on tests of distributions of closure quantities. In this approach we rely on the following modified median absolute deviation statistic

$$\text{mad}_0(Y) = 1.4826 \text{ med}(|Y|),$$

where ‘med’ denotes median, the subscript zero indicates that the raw distribution moment is estimated, and the normalization factor of 1.4826 scales the result so that it acts as a robust estimator of standard deviation for a normally distributed random variable $Y$ with zero mean.

We assume total uncertainties $\sigma$ associated with closure quantities to be well approximated by

$$\sigma^2 = \sigma^2_{\text{th}} + s^2,$$

such that the total uncertainty consists of the known a priori thermal component $\sigma_{\text{th}}$ and a constant systematic non-closing error $s$, of unknown magnitude, added in quadrature. We then solve for the characteristic value of $s$ that enforces

$$\text{mad}_0\left(\frac{X}{\sigma}\right) = \text{mad}_0\left(\frac{X}{\sqrt{\sigma^2_{\text{th}} + s^2}}\right) = 1,$$

149
where \( \sigma \) is the total uncertainty associated with \( X \). As an example, for RR–LL consistency of closure phases we have

\[
X = \psi_{C,RR} - \psi_{C,LL},
\]

\[
\sigma^2 = \sigma^2_{\psi_{C,RR}} + \sigma^2_{\psi_{C,LL}} + s^2.
\]

(5.25)

We exclude low S/N data (S/N < 7), for which the normal distribution approximation does not hold well.

**RR–LL Consistency**

Consistency of closure quantities derived from RR and LL visibilities, matched for the same scan, baseline and band, are expected to be dominated by effects related to polarimetric leakage, which remains uncalibrated in SR1 data. Assuming that some amount of leaked polarized signal mixes randomly into the parallel hand visibilities, the degree of systematic error can be crudely approximated as

\[
\delta_{\text{leak}} \approx \sqrt{2n|D||\tilde{m}|} < 0.14\sqrt{n}|\tilde{m}|,
\]

(5.26)

where the number of baselines \( n \) is 3 for closure phases and 4 for closure amplitudes, \( |D| < 0.1 \) is a leakage D-term magnitude and \( |\tilde{m}| \) is a typical fractional interferometric baseline polarization (i.e., fractional linearly-polarized correlated flux density relative to total intensity), see Johnson et al. [2015]. If a characteristic \( |\tilde{m}| < 0.2 \) is assumed, these upper bounds translate under Equation 5.26 to < 2.8° for the closure phase systematic uncertainty and < 5.7% for the closure amplitude uncertainty. The results of SR1 errors estimation by normalizing mad_0 are summarized in Table 5.7. The estimated errors are consistent with the simple upper limit given by Equation 5.26 and roughly consistent between all data reduction pipelines. While for the high S/N source 3C279 the leakage related errors may dominate over the thermal errors, they remain strongly subthermal for M87.
### Table 5.8: Inter-pipeline consistency of the SR1 data set.

<table>
<thead>
<tr>
<th></th>
<th>M87</th>
<th>3C279</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HOPS-</td>
<td>HOPS-</td>
</tr>
<tr>
<td></td>
<td>CASA</td>
<td>AIPS</td>
</tr>
<tr>
<td>Median visibility error (%)</td>
<td>3.8 (0.7)</td>
<td>7.9 (1.5)</td>
</tr>
<tr>
<td>90th percentile visibility error (%)</td>
<td>22.8 (6.0)</td>
<td>52.9 (7.4)</td>
</tr>
<tr>
<td>Median closure phase error (deg)</td>
<td>3.1 (0.3)</td>
<td>6.8 (0.7)</td>
</tr>
<tr>
<td>90th percentile closure phase error (deg)</td>
<td>17.7 (0.9)</td>
<td>39.4 (1.9)</td>
</tr>
<tr>
<td>Median log closure amplitude error</td>
<td>0.1 (0.3)</td>
<td>0.3 (0.9)</td>
</tr>
<tr>
<td>90th percentile log closure amplitude error</td>
<td>0.5 (1.0)</td>
<td>1.4 (2.4)</td>
</tr>
</tbody>
</table>

Note – Results given for scan-averaged single-band Stokes I data. Numbers in parentheses are given in thermal error units. The subset of data shared by all pipelines was used.
Frequency Bands Consistency

Comparisons between low/high frequency bands may reveal the presence of band-specific systematics, including frequency dependent polarimetric leakage. Apart from those, source spatial structure and spectral index both may add a small contribution. The estimated magnitudes of systematic errors found for closure phases and log closure amplitudes are given in Table 5.7. For all pipelines, the magnitude of characteristic closure phase inconsistency was found to be about 0.5 times the thermal uncertainty for M87 and about 1.5 times the thermal uncertainty for 3C279 (scan-average, single band/polarization). For 3C279 systematic uncertainties strongly dominate over the thermal scatter, and this should be taken into account before direct averaging of frequency bands.

Trivial Closure Quantities

The intra-site baselines ALMA–APEX and JCMT–SMA provide the EHT array with multiple ‘trivial’ closure triangles and quadrangles. Ideally, these trivial closure phases and trivial log closure amplitudes should be equal to zero, but this is not precisely true in the presence of polarimetric leakage. Furthermore, the small but finite length of intra-site baselines leads to measurements susceptible to contamination from large scale structure, breaking the assumptions of a trivial closure quantity. This particular aspect is a concern for M87 and its large scale jet. The estimated characteristic magnitude of systematic errors in trivial closure phases is given in Table 5.7. While for 3C279 the magnitude of about 1° can be fully explained by polarimetric leakage, M87 systematics are inconsistent with limits given by Equation 5.26, suggesting the presence of an additional source of error. We illustrate the systematic error fitting procedure in Figure 5.19, in which 3C279 trivial closure phases distribution is shown, before and after adding the systematics, estimated to be about 1° consistently for all processing pipelines.

Systematic Error Budget

Based on values reported in Table 5.7, we conclude that, for a single band, systematic errors of 3C279 measurements are dominated by polarimetric leakage and its contribution can be ap-
Figure 5.20: Closure statistics distributions after inflating errors by the amount of non-closing systematics recommended in Section 5.8.4. The plots follow the same order as the tests reported in Table 5.7. The dashed lines represent a standard normal distribution, and numbers show the fraction of $3\sigma$ outliers. Combined errors are used where appropriate.
proximated with characteristic values of about $1.5^\circ$ for closure phases and 0.03 for log closure amplitudes. For M87, leakage is not nearly as important, and other subtle effects like polarimetric calibration uncertainties may influence the total systematic error budget. Suggested systematics are $2^\circ$ for closure phases and 0.04 for log closure amplitudes. For each test of closure phases and log closure amplitudes summarized in Table 5.7, we show related distributions in Figure 5.20. Errors in Figure 5.20 were inflated according to the above recommendation for systematic errors. A standard (zero mean, unit variance) normal distribution is shown with a dashed line. The match between the empirical distributions and the normal distribution indicates that the addition of the systematic uncertainties allows to approximately capture the total data uncertainty. Under the assumption of independent baseline errors, the closure uncertainties given in this section can be translated to 2% non-closing systematic uncertainties in visibility amplitudes and 1° of non-closing systematic uncertainties in visibility phases.

5.8.5 Inter-Pipeline Consistency

Direct comparisons between corresponding data products delivered by separate pipelines allow us to quantify the degree of confidence we may have in their properties and their dependence on specific choices in calibration procedure. Figure 5.21 (top) shows the distribution of visibility amplitude differences between the reduction pipelines, in units of their thermal uncertainty. Thermal errors represent a particular scale of interest; however visibilities reduced by separate pipelines are not independent variables and share the same thermal noise realization. Another useful quantity is the relative absolute amplitude difference. As indicated in Table 5.8, the median relative difference between the most consistent pair of pipelines, HOPS–CASA, is 3.8%, well within the budget of a priori flux density calibration (Section 5.6). While for 3C279 all three pairs represent a similar level of consistency, for M87 the HOPS–CASA pair is by far the most consistent one, as indicated in Table 5.8. This result is consistent with known difficulties in the processing of low S/N data with the AIPS pipeline, originating from the lack of S/N to constrain a fringe solution in the two-second intervals used for fringe fitting (Section 5.5.3). Distributions of differences between amplitude data products are unbiased, however significant tails are present, with 10% of the M87 visibility amplitude data inconsistent by more than 22.8% for the most consistent pair, HOPS–CASA.

In Figure 5.22 we show HOPS–CASA and HOPS–AIPS scatter plots of correlation coefficient amplitude $|r_{ij}|$. The three pipelines demonstrate increasing levels of consistency at high S/N. AIPS shows a tendency to occasionally overestimate amplitude at low S/N, sometimes by a large factor, indicating a degree of over-tuning and acceptance of possible false fringes.

Contrary to visibility amplitudes, the distributions of closure phase and closure amplitude differences, shown in Figure 5.21, generally exhibit a spread at or below the level of thermal uncertainty, particularly for the HOPS–CASA pair. No significant tails are present and 90% of the M87 data remain consistent to within 0.9 standard deviations of the combined thermal error budget for HOPS–CASA (Table 5.8). This highlights the robustness of the closure quantities, independent of station-based gains.

Examples of closure phases for all three pipelines, for some of the triangles discussed in Section 5.7 are shown in Figure 5.23. While there is a broad consistency, HOPS is unique in reconstructing
Figure 5.21: Consistency of visibility amplitudes (top), closure phases (middle), and log closure amplitudes (bottom) between the three reduction pipelines. Scan-averaged single-band Stokes $I$ data are used.
Figure 5.22: Scatter plots of complex correlation coefficient amplitudes for HOPS–CASA and HOPS–AIPS pairs of pipelines. Data are fully averaged, with a S/N > 1 threshold applied. For each detection, the mean $r_{ij}$ of available RCP and LCP components in low and high band is given. Detections only present in one of the pipelines are shown with a fixed value of $5 \times 10^{-7}$ for the missing pipeline, and in some cases represent differences in the construction of a priori flags and fringe rejection strategies.

Figure 5.23: Comparison of M87 closure phases between the three fringe fitting pipelines for selected triangles. April 6 is shown in the top row, April 11 in the bottom row. The pipelines are offset slightly in time for clarity (HOPS -3min, CASA at the original timestamp, AIPS +3min). Plotted uncertainties denote $\pm 1\sigma$ ranges from thermal noise in the fully-averaged data set. For the two Hawaii triangles that demonstrate pronounced evolution on April 11 (see also Figure 5.14, bottom panels), we also include the corresponding redundant triangles with JCMT (which joined the array two scans earlier) as light crosses.
well-behaved closure phases on triangles including the LMT–SMA baseline over the full range of observations on April 11. To corroborate smooth trends and large closure phase evolution for these data, in two panels in Figure 5.23 we show data from a redundant JCMT triangle (JCMT and SMA are collocated). The redundant JCMT triangles show closure phases consistent with their SMA counterparts, and are more consistently reconstructed across the pipelines.

A bias towards zero closure phase can be seen when data are averaged in time, particularly for the AIPS data set. This is due to use of a point source model during global fringe fitting on short time intervals (2 s for AIPS). While the individual fringe solution phases are station-based and separately close, the process biases baseline phases to zero, and closure phases generated from baseline phases averaged over multiple segments will be biased toward the point-source model. This bias is not expected in HOPS products, as HOPS fringe solutions are baseline-based and assume no structure phase for the coherent stacking of data from multiple baselines. The median bias towards zero closure phase, estimated from high S/N data at least 3σ away from zero, is about 1° for AIPS and CASA with respect to unbiased HOPS. However, while 90% of CASA data are biased by less than 4.9°, 10% of AIPS data are biased by more than 8.7°. See Wielgus et al. [2019] for an additional discussion of pipeline comparisons and associated systematics.

The HOPS pipeline benefited from a long period of development, extensive review and internal validation through the suite of five engineering releases spanning a year-long data processing and calibration effort. In contrast, the AIPS pipeline has been used in two data releases as a secondary data set and the CASA pipeline, which is under active development, has recently been brought to maturity and included in ER5. Nonetheless, inter-pipeline comparisons of HOPS, CASA, and AIPS show a high degree of general consistency. The HOPS pipeline product was chosen as the primary scientific data set for SR1, based on the long validation history, level of calibration quality presented in this section, and to select a single data set for the preparation of scientific results. The other two pipelines are included in SR1 as supporting data sets for calibration, direct data comparisons, and as an independent pathway for validating the products of downstream analysis.

5.9 Conclusions

Observations from the Event Horizon Telescope’s April 2017 campaign are the first ever to have the necessary sensitivity, coverage, and resolution for horizon-scale imaging of black hole candidates M87 and Sgr A*. We have presented the complete data processing pathway that led to the first science release data set from the campaign, which includes the primary science target M87 and the secondary target 3C279. The 2017 observations reflected a dramatic expansion of the EHT from previous years to a total of eight sites, and include for the first time ALMA as a phased array. While much more powerful, the expanded network represented a unique analysis challenge in terms of the heterogeneous nature of the array: basic telescope characteristics, weather, sensitivity, site-specific data issues, sampling rate and channelization; and a challenge in terms of raw data volume and the needs for a homogeneous and systematized calibration strategy.

The development of processing pipelines and characterization of the data occurred over a series of five internal engineering releases, during which site-specific data issues were identified and mitigated in correlation and postprocessing. SR1 is the first science release of calibrated data products arising from the mature reduction pipelines, following a series of independent internal
reviews. The science data were produced without making assumptions about the detailed compact structure of the targets, and thus provide an unbiased data set for downstream imaging and modeling.

We have developed three independent processing pipelines for the initial fringe detection, phase calibration, and reduction of EHT data. The pipelines used HOPS, which has been continually developed and used for early EHT analysis over the previous decade; AIPS, the standard calibration environment for VLBI data from major facilities such as the VLBA; and CASA, a modern environment for radio interferometer calibration and analysis which has recently been augmented with VLBI capabilities. The output from each pipeline was subjected to a suite of validation tests covering self-consistency over bands and polarizations, and consistency of trivial closure quantities.

From these tests, we estimated the residual non-closing systematic errors after calibration. For M87 such errors remain smaller than Stokes $I$ data thermal uncertainties even after full scan and frequency band averaging. Non-closing errors are no larger than $2^\circ$ for closure phases and 4% for closure amplitudes. For 3C 279, systematics are small in an absolute sense, but they dominate the total uncertainties of the averaged data set due to the high S/N. Differences between pipelines, particularly for the robust closure quantities, were found to be largely within the total budget of uncertainties. The HOPS data were selected as the primary data set for the scientific conclusions presented in companion papers [Paper I; Paper IV; Paper V; Paper VI] with the remaining two data sets available for direct data comparisons and the cross-validation of downstream analysis.

At EHT frequencies, absolute flux density calibration is particularly challenging due to the large and time-varying 1.3 mm opacity from atmospheric water vapor, and difficulties maintaining pointing and surface accuracy particularly at the larger dishes. We have outlined the gathering and unified interpretation of auxiliary calibration data from the various sites for the purposes of a priori flux density calibration, and a strategy for estimating the residual flux density error budget within the limitations of single-dish calibration. Where available, we have made use of network redundancy to further constrain flux density calibration given generic model-independent assumptions about the source.

A number of salient features became apparent in the M87 data set after processing and calibration. The visibility amplitudes as a function of projected baseline length persistently show a prominent secondary peak bracketed by two nulls, the first at $\sim3.4 \, \text{G} \lambda$ and the second at $\sim8.3 \, \text{G} \lambda$, across all four observed days. The visibility amplitudes exhibit characteristics of a compact source with a spatial scale $\lesssim50 \, \mu\text{as}$, and broad circular symmetry broken on baselines probing the first null. This spatial scale corresponds to only a few Schwarzschild radii for a $\sim6.5 \times 10^3 \, M_{\odot}$ black hole [Paper VI] at the distance of M87 [Blakeslee et al., 2009; Gebhardt et al., 2011; Cantiello et al., 2018]. M87 closure phases on select triangles show clear time evolution between the two pairs of days, April 5/6 and April 10/11, providing evidence for intrinsic evolution of the source. The triangles with the largest closure phase variations between the two pairs of days have a baseline probing the $(u, v)$ plane region about the first minimum in visibility amplitude. Analysis and interpretation of these features are presented in companion papers [Paper I; Paper IV; Paper V; Paper VI].

Although previous observations of M87 from early EHT campaigns (in 2009 and 2012) probed scales of a few tens of micro-arcseconds, the visibility amplitude behavior on the few baselines
present remained consistent with a Gaussian source, showing no apparent finer structure [Doeleman et al., 2012; Akiyama et al., 2015]. The first M87 closure phases at 1.3 mm reported in Akiyama et al. [2015] were consistent with zero to within $2\sigma$. In addition to a first reported measurement of 1.3 mm closure amplitudes, the 2017 observations of M87 are the first to show non-Gaussian structure in the compact source and significantly non-zero closure phases.

The SR1 data provide the first opportunity for total intensity imaging of M87 [Paper IV]. Efforts to characterize and remove polarization leakage are ongoing and will enable studies of the linear polarization structure of M87 and other EHT targets. Additional work to better calibrate in the presence of intrinsic source variability, as well as increased amplitude gain variability is necessary for Sgr A* and other low-elevation targets.

For 2018, the EHT was joined by the Greenland Telescope, greatly expanding the coverage for northern sources such as M87. In the near future, the array will also be joined by the Kitt Peak 12-m telescope in Arizona and the NOEMA phased array at the Plateau de Bure observatory in France. In addition to generally improved baseline coverage, both sites provide short baselines and associated redundancy (with SMT and PV respectively) for the array – which is particularly beneficial for amplitude calibration. The EHT doubled recorded bandwidth to a rate of 64 Gbps in 2018 as well, over four 2 GHz bands. Additional development to enable coherent fringe fitting and atmospheric phase correction across all four bands will allow the EHT to better resolve features on long baselines, short timescales, and near visibility nulls, and it will increase robustness of the array against poor weather and the potential loss of sensitive central anchor stations.

While continuous development of the instrument and the data reduction pipeline will yield future observations with improved $(u,v)$ coverage, higher S/N and sharper resolution, the observations carried out in 2017 already deliver data of unprecedented scientific quality. The dramatic difference between the 2017 observations and early EHT campaigns in number of participating stations, S/N, coverage, and weather conditions make the EHT 2017 data set an exceptional opportunity for scientific discoveries via, e.g., imaging and model fitting well beyond previous EHT capabilities.

Acknowledgements

We wish to thank Eric W. Greisen for extending the capabilities of the AIPS software package, which made it possible to handle the wide-bandwidth EHT data. We acknowledge George Moellenbrock for his invaluable contribution to the CASA VLBI upgrade.

5.A Site and Data Issues

5.A.1 Issues Requiring Mitigation

The JCMT and SMA are located within hundreds of meters of each other on Mauna Kea. The small natural fringe rate is insufficient to wash out unwanted signals on the JCMT–SMA baselines (to phased and single-dish SMA). The JCMT and the SMA used identical frequency setups in 2017, resulting in two types of spurious correlations. For correlations between JCMT and the SMA single-dish reference antenna (not used directly for science analysis), two narrowband terrestrial
signals required special handling: one from the 1024 MHz spur tone of the R2DBEs, and a second one from the YIG oscillator tone (which is part of the LO chain) locally generated at the SMA. These signals were mitigated by flagging the affected frequency channels in post-processing.

Broadband celestial signals in the lower sideband with respect to the 220.1 GHz first LO used at the JCMT and SMA also contaminated the signal in the upper-sideband data. The differential fringe rate between upper and lower sidebands is of $O(\text{Hz})$; thus, the lower-sideband contamination averages out to zero over sufficiently long integration times. The contamination only affects the reference antenna contribution to the phased array as other antennas are subject to $90^\circ/270^\circ$ Walsh switching [Thompson et al., 2017, Section 7.5] that removes on average the lower sideband signal over a Walsh cycle of 0.65 s. Correlations between the JCMT and SMA single-dish reference antenna thus get the full lower sideband contribution, but correlations between JCMT and SMA phased array only get $1/N$ contribution, where $N$ is the number of telescopes being phased. To avoid phase steering toward this spurious $\sim 17\%$ contribution to the signal, neither the SMA nor the JCMT is ever used as the reference station during atmospheric phase calibration. For scans with very small fringe rates, there may be a small residual contribution after the 10 s averages used for network calibration (Section 5.6.2). This adds to the intra-site baseline amplitude error budget that propagates into gain solutions for that procedure, as well as for closure amplitudes that use the baseline on comparable timescales.

Data from PV were subject to substantial amplitude loss due to instabilities in the signal chain, attributed to excess phase noise in the maser frequency reference (which has since been replaced). Examination of the data on the ALMA–PV baseline with progressively shorter accumulation periods demonstrated a pattern of frequency spikes off the main signal with evidence that the full correlated amplitude could be recovered with an accumulation period of 2.048 ms. Further examination of a variety of scans showed that the pattern of frequency spikes was stable across scans, sources, and days, and the amplitude loss was constant. The effect was mitigated by continuing to use the data with a 0.4 s accumulation period and multiplying the visibility amplitudes on baselines to PV by a constant derived multiplicative factor of 1.914 during a priori flux density calibration, which is equivalent to multiplying the effective SEFD for PV by 3.663.

Misconfigured Mark6 recorders at APEX caused substantial data loss on many scans. The first 20–30 seconds of recording on a particular scan (sometimes much longer) were generally good, but partial or complete data dropouts could occur thereafter. DiFX accounts for the amount of valid data and automatically corrects averaged amplitudes and data weights for partial data loss to within $\sim 1\%$ accuracy. The remaining data from long-duration dropouts were manually flagged to avoid introducing bad APEX data into the processed data. The consequence is that ALMA–APEX coverage is inconsistent, and this complicates the strategy for network calibration and closure amplitude analysis, which makes use of intra-site baseline coverage. It also means that for the 2017 observations, APEX cannot be consistently used to help calibrate ALMA amplitude variation during poor weather when ALMA phasing efficiency is unstable.

A separate unrelated small correction factor is applied to APEX baselines to account for reduction in amplitude from the introduction of a 1 pulse-per-second signal (PPS) in the APEX data. The factor is estimated by measuring amplitudes with and without the PPS signal flagged. It is valid for multi-second averages of visibility amplitudes.

Isolated groups of frequency channels in the beamformer system at the SMA were occasionally
corrupted, causing a small fraction of the bandwidth (in the high band) to be lost during the first three days of the observation. Processing of a single band within the SMA beamformer is divided across eight hardware units, each of which processes one eighth of the total bandwidth, distributed across 128 channels of 2.234375 MHz each [Primiani et al., 2016], so that the exact pattern of lost channels, once identified, is predictable. The times when the data corruption occurred and the amount of bandwidth affected were identified using the strong noise correlation signal between the SMA (beamformed) phased array and the SMA single dish reference (recorded on a standard EHT backend). The pattern of lost bandwidth is evenly distributed throughout the band, and we derive SEFD corrections to account for the effective relative signal power lost upon frequency average (Table 5.2).

The LMT data are contaminated by polarization leakage which is delayed from the primary signal by \( \sim 1.5 \text{ns} \). This occurs in both polarizations, and is attributed to reflections in the optical setup of the LMT receiver used in 2017 (1.5 ns corresponds to 45 cm). The level of polarization leakage is \( \sim 10\% \), but for an unpolarized source it will dominate the correlated signal power of cross-hand VLBI products, therefore causing a false fringe at the delayed location. During fringe closure with the HOPS pipeline, an additional 1.5 ns delay systematic is added in quadrature to LMT baselines so that any such false fringes will not bias the global station delays. A future polarization leakage correction will need to accommodate leakage at non-zero delay to properly account for the contamination. For 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].

5.A.2 Issues Not Addressed During Processing

The failure of a hard drive in one of the JCMT modules caused one-sixteenth of the data in the low band to be lost. The lost data affects all scans on the module approximately equally, as packets are scattered onto all hard drives at record time. This issue required no special handling because DiFX automatically adjusts data weights based on the amount of data in each accumulation period.

Due to a small glitch in the ALMA correlator, the correlation coefficients on ALMA baselines are observed to undergo a slight dip every 18.192 s. The effective amplitude loss on scan-averaged quantities, less than 0.1\%, is well within the error budget and therefore unmitigated.

No corrections were made for losses due to finite Fast Fourier Transform (FFT) lengths, which are required to be long in order to align ALMA $32 \times 58.59375$ MHz data in the frequency domain with the wideband 2048 MHz single-channel data from most EHT stations. A small loss is introduced due to the changing delay over the 64 $\mu$s of time corresponding to the FFT length used. The loss is zero at the DC edge of the channel and increases linearly with frequency. This effect is baseline-dependent and greatest on the baselines with the greatest east-west extent, especially when the source is rising at one location and setting at the other. Across all fringes on all sources on all baselines on all five days, the median signal loss is 0.67\%, with the worst case (on a scan on the Hawaii–PV baseline) about an order of magnitude larger. FFT losses are negligible on baselines to ALMA because the delay error accumulates over a maximum of 58.59375 MHz in frequency rather than 2048 MHz.

The LMT faces significant challenges in maintaining an accurate surface for 1.3 mm as the
temperature fluctuates over the course of the evening. Pointing was also a challenge for scans at low or high elevation. These issues result in large residual gain trends obtained via amplitude self-calibration beyond the nominal error budget [Paper IV]. However, the station-based amplitude gain issues do not influence robust interferometric closure quantities.

The SPT, participating for the first time in the VLBI observations suffered from pointing problems early in the campaign. 3C279 observing time was used to diagnose and resolve these issues, resulting in missing a majority of 3C279 scans on April 5 and 6. The pointing issues were known and captured in observing logs during the run. The non-detections do not appear in the 3C279 data set (Figure 5.2), and their absence is expected.

5.A.3 Issues at Correlation

Two unanticipated issues with the ALMA data were discovered and fixed in a Rev7 correlation. First, the tuning of one of the ALMA local oscillator (LO) generators was specified to insufficient precision, resulting in an undocumented 50mHz LO offset. In most VLBI experiments, such a small LO offset might be transparently compensated by a small change in fitted delay-rate. However for the wide EHT bandwidths, the inability for a single delay-rate to model the effect over the entire 2GHz band is noticed, where the result of imperfect correction is to imprint a small rate slope with frequency, or, equivalently, a small delay drift with time. For this reason, the effect is separately corrected for prior to fringe fitting when postprocessing Rev5 data, which is possible for sufficiently small LO offsets.

Second, it was discovered that the ALMA delay system automatically removes the bulk atmospheric delay from above the array. By default, DIFX tries to remove the bulk atmospheric delay from above each station, resulting in a double correction for ALMA. This was most noticeable at low elevation, where the double correction imprinted a large and rapidly (but monotonically) changing delay rate. The large residual delay-rate is not large enough to cause decoherence over the duration of a correlation AP (0.4 seconds). The changing delay rate causes substantial decoherence over a several-minute scan if only a first-order fringe solution is used. Since EHT data reduction already includes a mechanism to measure and correct for non-linear phase due to atmospheric turbulence, it can also compensate for this drift in delay-rate imprinted on the data in the initial correlation. So long as signal-to-noise is sufficient to measure phase over short timescales, the impact on calibrated data is negligible.

Both of these issues were ultimately corrected in a final Rev7 correlation release. This included the LO adjustment for ALMA as well as special scripting for the geometric model preparation to allow the normal atmospheric correction at all sites other than ALMA merged with a no-atmospheric correction at ALMA. Comparison of SR1 results with comparable processing of Rev7 shows no significant difference, showing that the effects were sufficiently mitigated in postprocessing for SR1.
Chapter 6

Imaging a Black Hole with the EHT: Total-intensity and Polarimetric Imaging of M87*


Abstract

Efforts to image the supermassive black hole in the galaxy M87 with the Event Horizon Telescope (EHT) require a strong feedback connection with the data processing and calibration efforts. To build confidence on the source structure, exploration of software limitations and data-driven image properties are a crucial component of the imaging process.

This Chapter compiles the imaging and image validation efforts for the M87 total-intensity and polarimetric results published by the EHT. The validation procedures presented here, only a subset of the procedures for the final results, focus on our understanding of the antenna behaviors and the instrument as a whole. Comparisons are presented of M87 and calibrator amplitude gains and polarization leakage solutions across softwares and days. Connecting images and image-derived quantities to our understanding of the instrument, complementary to other validation tests, increases the fidelity and credibility of the M87 images in the wider very long baseline interferometry community.

163
Figure 6.1: (Top panels) Aggregate baseline coverage for EHT observations of M87, combining observations on all four days. The left panel shows short-baseline coverage, comprised of ALMA interferometer baselines and intra-site EHT baselines (SMA–JCMT and ALMA–APEX). These short baselines probe angular scales larger than 0.1″. The right panel shows long-baseline coverage, comprised of all inter-site EHT baselines. These long baselines span angular scales from 25 – 170 μas. Each point denotes a single scan, which range in duration from 4 to 7 minutes. (Bottom panels) The full baseline coverage on M87 for each observation. In all panels, the dashed circles show baseline lengths corresponding to the indicated fringe spacings (0.2″ for the upper-left panel; 25 and 50 μas for the remaining panels).
Figure 6.2: (Left) S/N as a function of projected baseline length for EHT observations of M87 on April 11. Each point denotes a visibility amplitude coherently averaged over a full scan (4-7 minutes). Points are colored by baseline. (Right) Visibility amplitudes (correlated flux density) as a function of projected baseline length after a priori and network calibration. The amplitudes are corrected for upward bias from thermal noise. Error bars denote ±1σ uncertainty from thermal noise and do not include expected uncertainties in the a priori calibration (see Event Horizon Telescope Collaboration et al. [2019c] and Section 6.1.1).

Figure 6.3: Selected closure phases from coherently averaged visibilities on three triangles as a function of GMST using data from all four days. Error bars denote ±1σ uncertainties from thermal noise. The trivial ALMA–APEX–SMT triangle (Left) has closure phases near zero on all days, as expected because this triangle includes an intra-site baseline. Deviations from zero arise from a combination of thermal and systematic errors [Event Horizon Telescope Collaboration et al., 2019c]. The ALMA–LMT–SMT triangle (Middle) shows persistent structure across all days, while the large LMT–SMA–SMT triangle (Right) shows source evolution between the first two days and last two days.
6.1 Total Intensity Imaging of M87

6.1.1 Observations and Data Processing

EHT Observations and Data

The EHT observed M87 with seven stations at five geographic sites on 2017 April 5, 6, 10, and 11. The participating facilities were the phased Atacama Large Millimeter/submillimeter Array (ALMA) and Atacama Pathfinder Experiment telescope (APEX) in the Atacama Desert in Chile, the James Clerk Maxwell Telescope (JCMT) and the phased Submillimeter Array (SMA) on Mauna Kea in Hawaii, the Arizona Radio Observatory Sub-Millimeter Telescope (SMT) on Mt. Graham in Arizona, the IRAM 30-m (PV) telescope on Pico Veleta in Spain, and the Large Millimeter Telescope Alfonso Serrano (LMT) on Sierra Negra in Mexico. These observations of M87 were interleaved with other targets (e.g., the quasar 3C279), some of which were visible to an eighth EHT station, the South Pole Telescope (SPT).

Data were recorded in two polarizations and two frequency bands. All sites except ALMA and the JCMT recorded dual circular polarization (RCP and LCP). ALMA recorded dual linear polarization later converted to a circular basis via \texttt{PolConvert} [Martí-Vidal et al., 2016b; Matthews et al., 2018; Goddi et al., 2019], and the JCMT recorded a single circular polarization (the recorded polarization varied from day-to-day).

All sites recorded two 2GHz bands centered on 227.1 and 229.1GHz (henceforth, low and high band respectively). Event Horizon Telescope Collaboration et al. [2019b] provides details on the setup, equipment, and station upgrades leading up to the 2017 observations.

Chapter 5 outlines the correlation, calibration, and validation of these data. In particular, the data reduction utilized the sensitive baselines to ALMA to estimate and correct for stable instrumental phase offsets, RCP–LCP delays, and stochastic phase variations within scans. After these corrections, the data have sufficient phase stability to coherently average over scans. The data were also amplitude-calibrated using station-specific measurements; stations with an intra-site partner (i.e., ALMA, APEX, SMA, and JCMT) were then “network calibrated” to further improve the amplitude calibration accuracy and stability via constraints among redundant baselines. The final network-calibrated data sets were frequency averaged per band and coherently averaged in 10-second intervals before being used for our imaging analysis. All data presented and analyzed in this Section are Stokes \( I \) (or pseudo \( I \)) visibilities processed via the EHT-HOPS pipeline (see Event Horizon Telescope Collaboration et al. [2019c] and Blackburn et al. [2019]).

Data Properties

Figure 6.1 shows the baseline \((u, v)\) coverage for EHT observations of M87. The shortest baselines in the EHT are intra-site (i.e., the SMA and JCMT are separated by 0.16 km; ALMA and APEX are separated by 2.6 km). These intra-site baselines are sensitive to arcsecond-scale structure. In contrast, our longest baselines (joining the SMA or JCMT to PV) are sensitive to microarcsecond-scale structure. Baseline coverage on individual days (bottom panels of Figure 6.1) is comparable

\footnote{Because the JCMT recorded a single circular polarization, baselines to JCMT use Stokes “pseudo \( I \).” Namely, we use parallel-hand visibilities to approximate Stokes \( I \) under the assumption that the source is weakly circularly polarized.}
for April 5, 6 and 11 (18, 25, and 22 scans, respectively). However, April 10 had significantly less coverage, with only 7 scans.

Figure 6.2 (left panel) shows the signal-to-noise ratio (S/N) as a function of baseline length for M87 on April 11, after coherently averaging scans. The split in S/N distributions at various baseline lengths is due to the sharp difference in sensitivity for the co-located Atacama sites ALMA and APEX. The right panel of Figure 6.2 shows the visibility amplitude (correlated flux density) for M87 on April 11 after amplitude and network calibration.

There is a prominent secondary peak in the network calibrated visibility amplitudes between two deep minima (“nulls”), the first at $\sim 3.4 G\lambda$ and the second at $\sim 8.3 G\lambda$. The amplitudes along the secondary peak are weakly dependent on baseline position angle, suggesting some degree of source symmetry, and their overall trends are consistent for all days (see Chapter 5, Figure 5.13). However, evidence for source anisotropy can be seen at the location of the first null, where the east-west oriented Hawaii–LMT baseline gives significantly lower amplitudes than the north-south oriented ALMA–LMT baseline at the same projected baseline length [see also Event Horizon Telescope Collaboration et al., 2019f]. This anisotropy is further supported by multiple measurements of non-zero closure phase (Figure 6.3). The majority of notable low-amplitude outliers across days are due to reduced performance of the JCMT or the LMT on a select number of scans. Despite the amplitudes of these data being low, the derived closure quantities remain stable [Event Horizon Telescope Collaboration et al., 2019c].

Similarly, most closure quantities for M87 are broadly consistent across all days, although day-to-day variations are significant for some sensitive closure combinations involving long baselines to PV or to the Hawaii stations. Figure 6.3 shows examples of closure phases for various triangles and levels of variability: a “trivial” triangle including co-located sites (ALMA–APEX–SMT, left panel) that is expected to be consistent with zero, a non-trivial and mostly non-variable triangle (ALMA–LMT–SMT, center panel) with largely persistent structure across all days, and a non-trivial triangle (LMT–SMA–SMT, right panel) showing intrinsic source structure evolution in M87 between the two sets of observations on April 5/6 and on April 10/11.

### Expected Amplitude Calibration Limitations

The amplitude calibration error budget is determined from uncertainties on individual measurements of station performance, as described in Chapter 5. The error budget is only representative of global station performance and is not specified for individual measurements. While this procedure is adequate for stations with stable performance and weather during the observing run, the error budget may be underestimated for stations with variable performance. The SMT, PV, SMA, JCMT, APEX, and ALMA stations are well-characterized either through years of studies or via network calibration. More recent additions to the EHT (the LMT and the SPT) may have more variable behavior, as their observing systems are not yet well exercised and because they do not have sufficiently close sites to permit network calibration.

Specifically for M87, the LMT is the most under-characterized station. The LMT began observing M87 in the evening, when the dish is still affected by thermal gradients in the back-up structure or panel distortions from day-time heating, both of which are significant for open-air telescopes. These effects are common for sensitive millimeter-wave dishes and cause surface
instability. In addition, evening conditions are inadequate for accurate pointing and focusing of
the telescope (particularly on weaker sources), leading to performance that can vary substantially
across scans and from day-to-day. A defocused dish can measure persistently low amplitudes on
baselines to that station between focus attempts (typically every one to two hours). Changes
in telescope pointing can cause amplitudes to fluctuate significantly from scan to scan (from the
telescope moving to and from the source) and between pointing attempts (typically every half-
hour). Issues in telescope focus can also lead to uncertainties in the a priori calibration for other
sources observed during the same time period, such as 3C 279. However, pointing errors for 3C 279
are expected to be less severe, as it is bright enough for the LMT to point directly on-source prior
to VLBI scans. Thus, the corrections needed for the LMT are expected to better match the a
priori amplitude error budget during observations of 3C 279 (mostly correcting for focus errors)
than during observations of M87 (correcting for both focus and pointing errors).

In Section 6.1.4 and Appendix 6.A, we compare estimated residual gains for the SMT and the
LMT from imaging M87 and 3C 279. In Event Horizon Telescope Collaboration et al. [2019f], we
compare these results with the estimated residual gains when fitting parametric models to the
interferometric data.

Image Conventions

Throughout this chapter, we present images using their equivalent brightness temperatures defined
by the Rayleigh-Jeans law: $T_b = \frac{c^2}{2\nu^2} I_{\nu}$, where $I_{\nu}$ is the specific intensity, $c$ is the speed of light,
$\nu$ is the observing frequency, and $k$ is the Boltzmann constant [e.g., Rybicki & Lightman, 1979].
We use brightness temperature rather than the standard radio convention of flux density per
beam (e.g., Jy/beam) because our images are spatially resolved and because RML methods do
not have a natural associated beam. However, we emphasize that brightness temperature does
not necessarily correspond to any physical temperature of the radio-emitting plasma. The radio
spectrum of M87 is not a blackbody, and its 230 GHz emission is from synchrotron radiation
[Event Horizon Telescope Collaboration et al., 2019e]. Finally, for visualization of our images, we
use perceptually uniform colormaps from the ehtplot\textsuperscript{2} library.

Throughout this chapter, inverse modeling images with CLEAN are convolved with a circular
Gaussian beam with 20 $\mu$as FWHM, comparable to the geometric mean of the principal axes of
the CLEAN beam. Any image restored with a beam will have the beam outlined in its lower right.
Also, for consistency with regularized maximum likelihood (RML) methods but in contrast with
standard practice, our presented CLEAN images do not include the residual image, corresponding
to the inverse Fourier transform of gridded residual visibilities. The characteristics of the residual
images are presented in Event Horizon Telescope Collaboration et al. [2019d] (hereafter Paper IV)
Section 7.2.

6.1.2 First M87 Images from Blind Imaging

VLBI images are sensitive to choices made in the imaging and self-calibration process. Choices
required in using any imaging method include deciding which data are used (e.g., low and/or high
band, flagging), specifying the self-calibration procedures, and fixing the reconstructed image

\textsuperscript{2}https://github.com/chanchikwan/ehtplot
FOV. In addition, imaging methods also require choices particular to their assumptions and methodology. For CLEAN, these choices include choosing a set of CLEAN windows and data weighting scheme. For RML methods, choices include the selection of which data and regularizer terms and weights to use in the objective function (see Chapter 3 Equation 3.14). With this abundance of user input, it can be difficult to assess what image properties are reliable from a given imaging method.

The dangers of false confidence and collective confirmation bias are magnified for the EHT because the array has fewer sites than typical VLBI arrays, there are no previous VLBI images of any source at 1.3 mm wavelength, and there are no comparable black hole images on event-horizon scales at any wavelength. To minimize the risk of collective bias influencing our final images, in our first stage of analysis we reconstructed images of M87 in four independent imaging teams.

Imaging Procedure and Team Structure

We subdivided our first M87 imaging efforts into four separate imaging teams. The teams were blind to each others’ work, prohibited from discussing their imaging results and even from discussing aspects of the data that might influence imaging (e.g., which stations or data might be of poor quality). No restrictions were imposed on the data preprocessing or imaging procedures used by each team. Teams 1 and 2 focused on RML methods, while Teams 3 and 4 primarily used CLEAN. I co-led the imaging efforts in Team 2 with Kazunori Akiyama. In addition to independently imaging M87, teams also independently imaged other sources observed by the EHT in 2017 to test the blind imaging procedure.

Blind imaging procedures have long been used to reduce the risk of group bias. Prior to the 2017 observations, we organized a series of “imaging challenges” that used synthetic data to assess how conventional and newly developed imaging algorithms would perform for the EHT [Bouman, 2017]. Reconstructing images independently in these challenges helped us identify which image features were likely intrinsic and which were likely spurious. To compare EHT 2017 results among teams while keeping submissions blind, we built a website that allowed users to independently upload images automatically compare them to the ground truth images and submissions from other users [Bouman, 2017].

Blind Imaging Within Team 2

Team 2 consisted of 16 members at various geographical locations, in the USA, in Japan and in the Netherlands. Team 2 was well balanced in terms of diversity of background, career stage and imaging software skill sets. Four software packages were used to image the ER4 data sets: the two RML methods SMILI (Akiyama et al. 2017a,b) and eht-imaging (Chael et al. 2016, 2018b); and more traditional CLEAN software packages such as the Common Astronomy Software Applications package (CASA; McMullin et al. 2007) and Difmap [Shepherd, 1997, 2011]. This enabled the team to apply various strategies and techniques, including newly developed imaging techniques such as maximum entropy, sparse modeling, dynamical imaging, and second moment imaging methods [e.g. Lawson et al., 2004].

---

3 Similar blind procedures have also been used in the optical interferometry community to evaluate and compare imaging methods [e.g. Lawson et al., 2004].
4 https://github.com/astrosmili/smili
5 https://github.com/achael/eht-imaging
regularization as well as traditional CLEAN techniques. Given the wide geographical coverage of the team membership, frequent communication and weekly work sessions and meetings were planned.

Before M87 data were made available to the imaging efforts, imaging challenges were carried out on the calibrator quasar sources J1924–2914 and 3C279. For both sources, members of the team carried out individual imaging with their preferred imaging software and technique, and cross-comparisons of image contours and derived station gain corrections were performed.

For J1924–2914, it was found that the compact flux density assumption produced different scaling in the station gain reconstructions (but consistent trends over time) but did not influence the image structure reconstructed, regardless of imaging technique or choices, see Figure 6.4. Comparisons of the image structure were also done with quasi-contemporaneous observations of J1924-2914 at 3.5mm [Chapter 3; Issaoun et al., 2019b], which showed a more extended jet structure in the north-west direction. We found consistency in the component direction between the 3.5mm and the 1.3mm images. Extended structure in the 1.3mm images was found to be unconstrained due to the dynamic range and field of view limitations of the EHT array. Final EHT images of J1924–2914 will be the subject of an upcoming publication (S. Issaoun et al. in prep.). The consistency between all images produced led to a unanimous team agreement that any image reconstructed can be considered a representative image to submit to the imaging challenge.

In the end, the reconstruction with SMILI where the intra-site baselines were removed was chosen as the representative submission from Team 2 for the imaging challenge because of its low noise level and improvement of the diversity of softwares and methods used across imaging teams.

For 3C279, the imaging of the data was more challenging. Convergence within the team was reached after a few iterations of individual imaging and comparisons. In the imaging process, it was found that using a 3.5 mm image of the source [Kim et al., 2020], to guide the jet direction of the blazar as a starting point helped convergence and the recovery of cleaner images with fewer spurious features. A mask was generated from the 3.5 mm image by blurring it with a 75 µas Gaussian beam and setting a threshold of 10% of the peak flux density. The mask was used as a prior in eht-imaging and as a strict field-of-view limit in SMILI. The source has a two to three-component core aligned North-West and a secondary component towards the south-south-west (Figure 6.5). The images show strong consistency among different methods for the peak positions of these components. These images of 3C279 were later used within the team to compare station gain corrections with M87 results.

Once M87 data were made available for the blind imaging stage, individual team members blindly imaged the 2017 April 11 data with no communication in the first week. Individual users were free to choose imaging parameters and prior information to input in the imaging process, yielding a wide variety of images that differed in total flux density, regularization method, field of view, usage of intra-site baseline information, and data product inputs, see Figure 6.6. While there are differences between individual reconstructions, all images show a dark spot in the brightness distribution. The representative team image for the 2017 April 11 data set was chosen from the SMILI reconstructions (purple border in Figure 6.6) to provide a software comparison across teams and due to the low level of spurious noise in the image. We compared the 3C279 station gains with M87 to check the consistency between them. As shown in Figure 6.7, the gains vary on hour timescales and show consistent trends between 3C279 and M87, although scatter and systematic
Figure 6.4: Self calibration amplitude gain trends for the 2017 April 11 J1924–2914 images reconstructed by Team 2. Consistent trends are found across individual reconstructions, but the overall gain level is sensitive to total image flux density assumptions by the user.
deviations are noticeable in some scans. These outlier scans are now understood to be caused by pointing errors at the LMT, see Appendix 6.A.

Following successful April 11 comparisons, we reconstructed M87 images individually for all epochs and different software choices. Examples of individual reconstructions are shown in Figure 6.8 with SMILI. We noted a shift of the peak emission from the south-east to the south portion of the ring between April 5/6 and April 10/11. This shift was later confirmed in the inter-team comparisons and in the results from the parameter surveys, see Section 6.1.3.

Additional tests were done within the team to understand the effects of field of view constraints, the inclusion of intra-site baseline information, and data variability on image morphology. We have concluded that limiting the field of view of the images lowers the noise level and avoids the intrusion of spurious structure. For M87, the optimal field of view was found to be around 80-100µas. Furthermore, noise features are reduced strongly when intra-site baselines are removed before imaging with visibility amplitudes. This is especially useful for sources with faint extended structure that is not sampled nor constrained by the current EHT coverage, the case for M87.

During discussions with the imaging effort at large, some concerns arose about the closure phase error budget, and the difficulty of producing images with a closure phase reduced $\chi^2$ lower than 2. In Team 2, we decided to especially pay close attention to fitting closure phases from large triangles, which show the most noticeable variations from day to day. With dynamical imaging [Johnson et al., 2017] across the campaign, it was found that since most triangles show very stable closure phases, dynamical reconstructions smooth out small structural variability on large triangles to produce a stable image of M87 throughout all days. However, with static day-to-day imaging and closer attention to fitting these large triangles, we noticed deviations in the images based on the fits to those triangles. All day-to-day reconstructions (both in SMILI and eht-imaging) showed a clear split in fits to the two first days and two last days of the observing run, where the pairs of reconstructions nearest in time fit closure phases very well for their corresponding days but do not fit the measurements on the other pair of days, as shown in Fig. 6.9 for our SMILI representative images. This is indicative of clear time variability signatures, notably changes in asymmetry probed by closure phases, over the course of the campaign, with
6.1 Total Intensity Imaging of M87

Figure 6.6: Individual images from team members after one week of blind imaging with no intra-
team communication. Four software packages were used: the RML softwares SMILI and eht-imaging, and the CLEAN softwares CASA and Difmap. Each column shows a reconstruction with a given software and user-based choices. Some differences in the choices are as follows: (1) Images using only VLBI data (no intra-site baseline information); (2) images using the entire data set; (A) imaging with closure quantities only; (B) imaging with visibility amplitudes and closure phases; (C) imaging with full complex visibilities. Each row shows the reconstructions with a different scale, to show the differences in noise level between reconstructions: linear scale (top), square-root scale (middle), and log scale (bottom). The reconstruction chosen as the representative team submission is marked by a purple border.

Figure 6.7: Gain plots of LMT (left) and SMA (right) for 2017 April 11. 3C279 and M87 are represented with pink and green points, respectively. Consistent trends between sources are seen (although the overall gain level depends on assumptions on total image flux density) while some outliers are indicative of telescope miscalibration. Gain trends are studied in the wider imaging effort in Section 6.1.4.
Figure 6.8: M87 reconstructions for each day (left to right) made with SMILI by individual members of the team (each row). While some features of the ring change between individual imaging choices, a shift in the brightness distribution from the south-east to the south of the ring from April 5 to April 11 is noticeable for all reconstructions.
the 2017 April 5/6 pair showing consistent structure but with notable differences from the April 10/11 pair.

![Figure 6.9: Closure phases on the largest triangle (ALMA–LMT–SMA) for different observing nights (2017 April 5, 6, 10 and 11). Each panel corresponds to the observed closure phases for a particular observing night and the fits to all four images from day-to-day static imaging of M87.](image)

**First M87 Imaging Results**

The first M87 imaging analysis used an early-release engineering data set [ER4; Event Horizon Telescope Collaboration et al., 2019c]. These data had a priori and network calibration applied but did not have calibrated relative RCP-LCP gains. Consequently, each team imaged the data using only parallel-hand products (i.e., RCP-RCP or LCP-LCP) to approximate total intensity. The April 11 data set was selected for the first comparison, since it had the best coverage for the M87/3C 279 pair and the most stable a priori amplitude calibration (especially for the LMT).

The imaging teams worked on the data independently, without communication, for seven weeks, after which teams submitted images to the image comparison website using LCP data (because the JCMT recorded LCP on April 11). After ensuring image consistency through a variety of blind metrics (including normalized cross-correlation, see Chapter 3 Equation 3.4.2), we compared the independently reconstructed images from the four teams.

Figure 6.10 shows these first four images of M87. All four images show an asymmetric ring structure. For both RML teams and both CLEAN teams, the ring has a diameter of approximately 40 μas, with brighter emission in the south. In contrast, the ring azimuthal profile, thickness, and brightness varies substantially among the images. Some of these differences are attributable
to different assumptions about the total compact flux density and systematic uncertainties (see Table 6.1).

The initial blind imaging stage indicated that the image of M87 is dominated by a \( \sim 40 \mu\text{as} \) ring. The ring persists across the imaging methods. Next, we moved to a second, non-blind imaging stage that focused on exploring the space of acceptable images for each method. The independent team structure was only used for the first stage of imaging; the remainder of this paper will categorize results by imaging methodology.

### 6.1.3 Parameter Surveys and Final Images

Following the preliminary imaging, a systematic approach was taken to test the three imaging pipelines used. While the first stage involved user-based decisions, a ranking system was devised to select optimal parameters for each pipeline that reconstructed a range of image structures where the ground truth is known with high fidelity. Synthetic data sets were created that mimic the visibility amplitude profile of M87 but with different underlying source structure, see Figure 6.11.

Four underlying source structures were selected: a uniform ring, an asymmetric ring crescent, a uniform disk, and a double source.

Parameter surveys were designed for the three software pipelines: the inverse modeling software Difmap [Shepherd, 1997, 2011] and the two regularized maximum likelihood (RML) forward modeling softwares eht-imaging [Chael et al., 2016, 2018b] and SMILI [Akiyama et al., 2017a,b]. Each parameter survey explored a range of parameter combinations to determine imaging fidelity of each combination against all synthetic data sets. The fiducial parameters were selected from a top set of images per pipeline that obtained high image fidelity in terms of image domain cross-correlation between the produced and the original image and least-squares (\( \chi^2 \)) fits to data.
### Image Properties

<table>
<thead>
<tr>
<th>Method</th>
<th>Team 1</th>
<th>Team 2</th>
<th>Team 3</th>
<th>Team 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{\text{cpct}}$ (Jy)</td>
<td>0.94</td>
<td>0.43</td>
<td>0.42</td>
<td>0.42</td>
</tr>
</tbody>
</table>

**Engineering Data (10-sec avg., LCP, 0% sys. error)**

| $\chi^2_{\text{CP}}$ | 2.06 | 2.48 | 2.44 | 2.33 |
| $\chi^2_{\log \text{CA}}$ | 1.20 | 2.16 | 2.15 | 1.43 |

**Science Release (scan-avg., Stokes $I$, 0% sys. error)**

| $\chi^2_{\text{CP}}$ | 1.13 | 5.40 | 2.28 | 1.89 |
| $\chi^2_{\log \text{CA}}$ | 2.12 | 5.41 | 3.90 | 5.32 |

**Science Release (scan-avg., Stokes $I$, 1% sys. error)**

| $\chi^2_{\text{CP}}$ | 1.00 | 3.85 | 2.04 | 1.55 |
| $\chi^2_{\log \text{CA}}$ | 1.96 | 5.07 | 3.64 | 4.8 |

**Science Release (scan-avg., Stokes $I$, 10% sys. error)**

| $\chi^2_{\text{CP}}$ | 0.49 | 0.95 | 1.11 | 0.48 |
| $\chi^2_{\log \text{CA}}$ | 0.46 | 1.36 | 0.98 | 0.79 |

Table 6.1: Image properties and data consistency metrics for the first M87 images (see Figure 6.10). Data metrics are shown as originally computed on April 11 data (using 10-second averaged engineering data with LCP) and using the data from the first EHT Science Release (scan-averaged, Stokes $I$) when 0%, 1% and 10% systematic error has been included. Teams 2–4 chose to exclude the intra-site baselines in their imaging. However, for consistency with our later $\chi^2$ values computed from Science Release data, we include these baselines when computing $\chi^2$ after adding an extended component to these images containing the missing flux density.
Figure 6.11: The four simple geometric models and synthetic data sets used in the parameter surveys. (Top) Linear scale images, highlighting the compact structure of the models. (Middle) Logarithmic scale images, highlighting the larger-scale jet added to each model image. (Bottom) One realization of simulated visibility amplitudes corresponding to the April 11 observations of M87. We indicate the conventions for cardinal direction and position angle used throughout this paper on the upper-right panel. Note that east is oriented to the left, and position angles are defined east of north. Figure from Paper IV.
Figure 6.12: Fiducial images of M87 on all four observed days from each of the three imaging pipelines. CLEAN images (from DIFMAP) are shown after convolution with a 20 µas beam; eht-imaging and SMILI results have no restoring beam applied. Different selected fiducial imaging parameters (e.g., compact flux) result in different peak brightness temperatures for each method, as indicated by the unique color bars for each row.
Figure 6.13: Fiducial images of M87 on April 11 from our three separate imaging pipelines after restoring each to an equivalent resolution. The \textit{eht-imaging} and SMILI images have been restored with 17.1 and 18.6\,\mu as FWHM Gaussian beams, respectively, to match the resolution of the DIFMAP reconstruction restored with a 20\,\mu as beam.

Figure 6.14: Averages of the three fiducial images of M87 for each of the four observed days after restoring each to an equivalent resolution, as in Figure 6.13. The indicated beam is 20\,\mu as (i.e., that of DIFMAP, which is always the largest of the three individual beams).

products. A detailed account of the parameter surveys is presented in Section 6 of Paper IV. While the entire top set for each pipeline was used for M87 image analysis, a single parameter combination was chosen as the ‘fiducial’ combination to produce a final image. The fiducial images produced at the pipeline native resolution (super-resolved for RML) are presented in Figure 6.12.

A level of blurring was calculated across softwares to maximize cross-correlation of the fiducial images: the native CLEAN resolution of 20\,\mu as was used for Difmap, while \textit{eht-imaging} and SMILI yielded a convolving beam of 17.1\,\mu as and 18.6\,\mu as respectively. The difference in effective blurring is due to the original compactness of the structure produced by each software (CLEAN produces an array of point-sources, thus has the highest level of blurring). The final beam-convolved images for 2017 April 11 are presented in Figure 6.13. This representation is the most conservative view of our image structure in M87, and differences between softwares are minimal. The final M87 images for each day are the averages of the three fiducial images across software pipelines, shown in Figure 6.14. The source structure for all four days is an asymmetric ring brighter in the south-south-east portion. Extraction of the ring size and studies of the variability of the brightness distribution of the ring across days are the subject of Section 9 of Paper IV.
6.1 Total Intensity Imaging of M87

6.1.4 Image Validation with 3C 279

Having determined fiducial images of M87 from each imaging method on each observing day, we now perform additional validation tests to assess their reliability. In this Section, we compare the residual telescope gains estimated for M87 with those for the calibrator source 3C 279 to determine whether the significant variation seen in the inferred LMT residual gains are consistent between the two sources. Additional validation tests of image morphology with data sampling and validation tests of time variability across days linked to \((u,v)\) coverage differences are presented in Sections 8.2 and 8.3 of Paper IV.

The 2017 EHT observations of M87 were interleaved with those of the active galactic nucleus (AGN) 3C 279. We have reconstructed images of 3C 279 using the three software packages used for M87 imaging in order to assess the consistency of these 3C 279 images with the M87 reconstructions in terms of the inferred time-variable residual gains. Since the sources are nearby on the sky (separation \(= 19^{\circ}\)), the inferred time-variable residual gains at each site for the same day should be similar. We do not use the observations of 3C 279 to derive gain corrections for M87; instead, we compare the derived gains on both sources after independent imaging as a post hoc consistency test.

Figure 6.15 shows the aggregate baseline coverage for the interleaved EHT observations of 3C 279 in April 2017. While the SPT could not observe M87, it participated in the observations of 3C 279, viewing it at an elevation of \(\sim 6^{\circ}\) (i.e., a relative air mass of \(\sim 10\)). The addition of the SPT significantly improves the north-south resolution of the array. Figure 6.16 shows the April 11 visibility amplitudes from 3C 279 after a priori and network calibration. From ALMA interferometric measurements, the total flux density for 3C 279 (8–10 Jy) is nearly an order of magnitude higher than that of M87 [Goddi et al., 2019]. As expected for a source with bright, linear jet features, the 3C 279 amplitudes vary strongly with baseline position angle and have a more complex structure on long baselines than M87 (Figure 6.2).

We imaged 3C 279 using both traditional CLEAN (DIFMAP) and RML (eht-imaging and SMILI) methods. Because of pronounced differences between the sources in different characteristics (total flux density, field of view, compact structure morphology), we do not use the same fiducial scripts derived for M87 on the 3C 279 data. Furthermore, as a consequence of having very few short baselines in the EHT array, we have found imaging 3C 279 to be more difficult than M87 because of its more extended structure. A full description of 3C 279 imaging procedures and results will be presented separately.

Figure 6.17 shows reconstructed images of 3C 279 from all three methods using data from April 11. The source exhibits two compact and bright features with a separation of \(\sim 100 \mu\text{as}\): a primary component extended in the northwest to southeast direction, and a secondary component perpendicular to the first. This secondary component extends from the core in the direction of the 3C 279 jet observed at lower frequencies [see, e.g., Lister et al., 2016; Jorstad et al., 2017].

In Figure 6.18, we compare the interleaved multiplicative station gains for M87 and 3C 279 on April 5. The station gains were derived via self-calibration to the fiducial images of M87 from the three imaging pipelines (Section 6.1.3) and on the three images of 3C 279 (Figure 6.17). In Table 6.2, we present the median gains for the two sources and compare them to the expected a priori visibility amplitude error budget, assuming nominal pointing and focus [Event Horizon
Figure 6.15: Aggregate baseline coverage for EHT observations of 3C 279 in April 2017. The coverage of M87 is shown in light gray for comparison.

Figure 6.16: Visibility amplitudes of 3C 279 on April 11 as a function of projected baseline length after a priori and network calibration. The amplitudes are corrected for upward bias from thermal noise. Points are colored by baseline as in Figure 6.15.
6.1 Total Intensity Imaging of M87

**Figure 6.17:** Representative images of 3C279 from the April 11 EHT observations produced using DIFMAP, eht-imaging, and SMILI. To simplify visual comparisons and display the images at similar resolutions, the images are restored with circular Gaussian beams of 20, 17.1 and 18.6 \( \mu \text{as} \) FWHM, respectively.

<table>
<thead>
<tr>
<th>Station</th>
<th>Fiducial M87 median gain</th>
<th>3C279 median gain</th>
<th>A priori budget (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIFMAP</td>
<td>eht-imaging</td>
<td>SMILI</td>
</tr>
<tr>
<td>ALMA</td>
<td>0.97^{+0.02}_{-0.03}</td>
<td>0.97^{+0.02}_{-0.01}</td>
<td>0.98^{+0.01}_{-0.01}</td>
</tr>
<tr>
<td>APEX</td>
<td>1.05^{+0.04}_{-0.02}</td>
<td>1.02^{+0.02}_{-0.01}</td>
<td>1.01^{+0.01}_{-0.01}</td>
</tr>
<tr>
<td>SMT</td>
<td>1.13^{+0.01}_{-0.06}</td>
<td>1.02^{+0.04}_{-0.01}</td>
<td>0.99^{+0.02}_{-0.01}</td>
</tr>
<tr>
<td>JCMT</td>
<td>1.00^{+0.02}_{-0.00}</td>
<td>1.00^{+0.00}_{-0.00}</td>
<td>1.00^{+0.00}_{-0.00}</td>
</tr>
<tr>
<td>LMT</td>
<td>1.46^{+0.03}_{-0.21}</td>
<td>1.55^{+0.19}_{-0.19}</td>
<td>1.47^{+0.13}_{-0.04}</td>
</tr>
<tr>
<td>SMA</td>
<td>1.00^{+0.02}_{-0.01}</td>
<td>1.00^{+0.00}_{-0.00}</td>
<td>1.00^{+0.00}_{-0.00}</td>
</tr>
<tr>
<td>PV</td>
<td>1.14^{+0.04}_{-0.05}</td>
<td>0.96^{+0.02}_{-0.07}</td>
<td>0.98^{+0.02}_{-0.07}</td>
</tr>
</tbody>
</table>

**Table 6.2:** Median, 25th, and 75th percentile residual gain corrections for M87 and 3C279 on April 5. These gains were derived via self-calibration (with no systematic error included). The error budget on a priori calibration is derived in Chapter 5. Note that the median gain corrections for ALMA, APEX, SMA, and JCMT can reasonably be much smaller than this error budget because network calibration has already been applied. The variation in the recovered gains among pipelines is partly due to the large uncertainty in the total flux density (between 8–10 Jy) and total compact flux density of 3C279.
Figure 6.18: Multiplicative residual station gains for the SMT (left) and LMT (right) derived from the 3C279 images (Figure 6.17) and fiducial M87 images (Figure 6.13) from the three imaging pipelines on April 5. Gains for the fiducial images of M87 are shown in red; those for 3C279 are shown in blue. The particularly large excursions on the LMT M87 gains are likely due to poor pointing. Note that LMT could not observe 3C279 before 2h30 UTC.

Telescope Collaboration et al., 2019c.

The gains derived from images produced with the three imaging methods are consistent. For all stations except the LMT, the derived gains are time-variable, but close to unity. Issues with the performance of the LMT (Section 6.1.1, and Event Horizon Telescope Collaboration et al. [2019c]) result in a larger median gain correction at that station, in line with the gain corrections derived directly from the visibility amplitudes via the use of crossing tracks on M87 described in Appendix B of Paper IV. Furthermore, the interleaved LMT gain curve has large variations with time on both sources. In particular, the M87 gains at the LMT have large, single-scan excursions not seen in the 3C279 gain trends. These excursions are likely due to poor pointing on M87, which is nearly an order of magnitude fainter than 3C279.

In Table 6.5 of Appendix 6.A, we present the median gains for the two sources, compared to the expected a priori visibility amplitude error budget, across all days. Figure 6.28 in Appendix 6.A compares the gain variations recovered from imaging M87 and 3C279 across all days for the SMT and LMT stations. In some cases, the absolute gain is not identical across the different sources and imaging pipelines. This variation can be partly ascribed to the large uncertainty in the total flux density of 3C279. However, in all cases the relative gain trends with time are broadly consistent, except for the occasional large excursions seen in the LMT M87 gains.

The consistency of the derived gain variations between the two sources and the different imaging methods for all stations on all days, particularly for the large corrections at LMT, provides confidence that these corrections are not the result of imaging artifacts or missing structure in M87 reconstructions. In Event Horizon Telescope Collaboration et al. [2019f] we show that derived gains from fitting parametric models are similar to those derived from imaging.
6.2 Polarimetric Imaging of M87

6.2.1 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, \( I(\mathbf{x}) \) (total intensity), \( Q(\mathbf{x}) \) (difference in horizontal and vertical linear polarization), \( U(\mathbf{x}) \) (difference in linear polarization at 45 deg and −45 deg position angle), and \( V(\mathbf{x}) \) (circular polarization). We define the complex linear polarization \( p \) as

\[
p \equiv Q + iU = I|m|e^{2i\chi},
\]

where \( m = (Q + iU)/I \) 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 \( \hat{I} \) as a function of the spatial frequency \( u = (u, v) \) of the total intensity image; similarly, polarimetric VLBI observations also sample the Fourier transform of the other Stokes parameters \( \hat{Q}, \hat{U}, \hat{V} \).

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

\[
\rho_{jk} \equiv \begin{pmatrix} R_j R_k^* & R_j L_k^* \\ L_j R_k^* & L_j L_k^* \end{pmatrix} = \begin{pmatrix} \hat{I}_{jk} + \hat{V}_{jk} & \hat{Q}_{jk} + i\hat{U}_{jk} \\ \hat{Q}_{jk} - i\hat{U}_{jk} & \hat{I}_{jk} - \hat{V}_{jk} \end{pmatrix},
\]

(6.2)

for a baseline between two stations \( j \) and \( k \). The notation \( R_j L_k^* \) indicates the 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 \( R_j \) measured by the telescope \( j \) and the left-hand circularly polarized component \( L_k \) measured by the telescope \( k \).

Equation 6.2 defines the coherency matrix \( \rho_{jk} \). Following Johnson et al. [2015], we also define the fractional polarization in the visibility domain,

\[
\tilde{m} \equiv \frac{\hat{Q} + i\hat{U}}{\hat{I}} = \frac{2RL^*}{RR^* + LL^*},
\]

(6.3)

Note that Equation 6.3 implies that \( \tilde{m}(u) \) and \( \tilde{m}(-u) \) constitute independent measurements for \( u \neq 0 \). Moreover, \( \tilde{m}(u) \) and \( m(x) \) 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 \( \tilde{m} \). Useful relationships between \( \tilde{m} \) 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 \( \Phi \) [Thompson et al., 2017],

\[
\mathbf{J} = \mathbf{G} \mathbf{D} \Phi = \begin{pmatrix} G_R & 0 \\ 0 & G_L \end{pmatrix} \begin{pmatrix} 1 & D_R \\ D_L & 1 \end{pmatrix} \begin{pmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix},
\]

(6.4)
Time-dependent field rotation matrices $\Phi \equiv \Phi(t)$ are known a priori, with the field rotation angle $\phi(t)$ dependent on the source’s elevation $\theta_{el}(t)$ and parallactic angle $\psi_{par}(t)$. The angle $\phi$ takes the form

$$\phi = f_{el}\theta_{el} + f_{par}\psi_{par} + \phi_{off},$$

(6.5)

where $\phi_{off}$ is a constant offset, and the coefficients $f_{el}$ and $f_{par}$ are specific to the receiver position type. The gain matrices $G$, containing complex station gains $G_R$ and $G_L$, are estimated within the EHT’s upstream calibration and total intensity imaging pipeline, see Section 6.2.2. Estimation of the D-terms, the complex coefficients $D_R$ and $D_L$ of the leakage matrix $D$, 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 6.2.3.

For a pair of VLBI stations $j$ and $k$ the measured coherency matrix $\rho'_{jk}$ is related to the true source coherency matrix $\rho_{jk}$ via the Radio Interferometer Measurement Equation, hereafter RIME [Hamaker et al., 1996; Smirnov, 2011],

$$\rho'_{jk} = J_j \rho_{jk} J_k^\dagger,$$

(6.6)

where the dagger $\dagger$ symbol denotes conjugate transposition. Once the Jones matrices for the stations $j$ and $k$ are well characterized, Equation 6.6 can be inverted to give the source coherency matrix $\rho_{jk}$. From $\rho_{jk}$, Stokes visibilities can be obtained by inverting Equation 6.2:

$$\begin{pmatrix}
\tilde{I}_{jk} \\
\tilde{Q}_{jk} \\
\tilde{U}_{jk} \\
\tilde{V}_{jk}
\end{pmatrix} = \frac{1}{2} \begin{pmatrix}
R_jR_k^* + L_jL_k^* \\
R_jL_k^* + L_jR_k^* \\
-\text{i}(R_jL_k^* - L_jR_k^*) \\
R_jR_k^* - L_jL_k^*
\end{pmatrix}.
$$

(6.7)

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

6.2.2 EHT 2017 Polarimetric Data

Correlation and Data Calibration

An overview of the EHT April 2017 observations was given in Section 6.1.1. 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 0.5MHz. 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 [Event Horizon Telescope Collaboration et al., 2019c]. 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
### 6.2 Polarimetric Imaging of M87

Table 6.3: Field rotation parameters for the EHT stations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Receiver location</th>
<th>$f_{\text{par}}$</th>
<th>$f_{\text{el}}$</th>
<th>$\phi_{\text{off}}$(deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA</td>
<td>Cassegrain</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>APEX</td>
<td>Nasmyth-Right</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>JCMT</td>
<td>Cassegrain</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SMA</td>
<td>Nasmyth-Left</td>
<td>1</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td>LMT</td>
<td>Nasmyth-Left</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>SMT</td>
<td>Nasmyth-Right</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PV</td>
<td>Nasmyth-Left</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>SPT</td>
<td>Cassegrain</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Astronomy Software Applications package [CASA; McMullin et al., 2007; Janssen et al., 2019b]; and the NRAO Astronomical Image Processing System [AIPS; Greisen, 2003; Event Horizon Telescope Collaboration et al., 2019c]. 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 6.4-6.5, using coefficients given in Table 6.3. Flux density (amplitude) calibration is applied via a common post-processing framework for all pipelines [Blackburn et al., 2019; Event Horizon Telescope Collaboration et al., 2019c], taking into account estimated station sensitivities [Appendix A; Issaoun et al., 2017a; Janssen et al., 2019a]. Under the assumption of zero circular polarization of the primary (solar system) calibrator sources, 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 $G$ 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 $\tilde{I}, \tilde{Q}, \tilde{U}, \tilde{V}$ 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 $I$ brightness distribution, preserving the complex polarimetric gain ratios (e.g., Event Horizon Telescope Collaboration et al. [2019d,f]). Fully calibrating the D-terms requires modeling the polarized emission.

The Stokes $I$ (total intensity) analysis of a subset of the 2017 observations (Science Release 1; SR1), including M87, was the subject of Event Horizon Telescope Collaboration et al. [2019a,b,c,d,e,f]. The quality of these Stokes $I$ 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 Chapter 5. Information about accessing SR1 data and the software used for analysis can be found on the Event Horizon Telescope website’s data portal.\footnote{\url{https://eventhorizontelescope.org/for-astronomers/data}} 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 [\citet{Martin-Vidal16,Godd19}]. Other minor subtleties in the handling of polarimetric data are presented in Appendix 6.B.

\section*{Polarimetric Data Properties}

In Figure 6.19 (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 \(\tilde{m}\) [i.e., the fractional polarization in visibility space; for analysis of \(\tilde{m}\) in another source, Sgr A*, see \citet{Johnson15}]. M87 is weakly polarized on most baselines, \(|\tilde{m}| \lesssim 0.5\). Several data points on SMA–SMT baselines have very high fractional polarization \(|\tilde{m}(u, v)| \sim 2\) that occur at \((u, v)\) spacings where the Stokes \(I\) visibility amplitude enters a deep minimum. The
fractional polarization $\tilde{n}$ 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 6.19 (lower panels) we show the field rotation angles $\phi$ 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 deg 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 for M87 in total intensity, 3C 279 [Event Horizon Telescope Collaboration et al., 2019d; Kim et al., 2020], is bright and strongly polarized on longer baselines and is not used in this work. The properties and analysis of the calibrators are discussed in more detail in Appendix 6.C.

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

6.2.3 Methods for Polarimetric Imaging and Leakage Calibration

Producing an image of the linearly polarized emission requires both solving for the sky distribution of Stokes parameters $Q$ and $U$ and for the instrumental polarization of the antennas in the EHT array. 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 of Event Horizon Telescope Collaboration et al. [2021a], hereafter Paper VII.

The calibration of the instrumental polarization by sub-component fitting was performed using three different codes (LPCAL, GPCAL, and polsolve) that depend on two standard software packages for interferometric data analysis: AIPS\textsuperscript{7} and CASA\textsuperscript{8}. In all of these methods, the Stokes $I$ imaging step is performed using the CLEAN algorithm [Högbom, 1974], and sub-components with constant complex fractional polarization are then 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\textsuperscript{9} [Park et al., 2021]. In CASA, we use the polsolve algorithm [Martí-Vidal et al., 2021], which uses data from multiple calibrator sources simultaneously to fit polarimetric sub-

\textsuperscript{7}http://www.aips.nrao.edu

\textsuperscript{8}https://casa.nrao.edu

\textsuperscript{9}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. For the M87 work, GPCAL is used to complement the LPCAL analysis of the M87 data and the D-term estimation using calibrators (Appendix 6.C). We do not show GPCAL results in the main text.
components and allows for D-terms to be frequency-dependent (see Appendix D of Paper VII). In all sub-component fitting and imaging methods we assume that Stokes $\nu = 0$.

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 230GHz core in M87. RML algorithms find an image that maximizes an objective function composed of a likelihood term and regularizer terms that penalize or favor certain image features. In this work, we use the RML method implemented in the eht-imaging\footnote{https://github.com/achael/eht-imaging} software library [Chael et al., 2016, 2018b] to solve for images in both total intensity and linear polarization. Like the CLEAN-based methods, eht-imaging does not solve for Stokes $\nu$. Details on the specific imaging methods in eht-imaging used in the reconstructions presented in this work can be found in Appendix C of Paper VII.

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 $\nu$) 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 of Paper VII and in separate publications (Pesce 2021, 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.

### 6.2.4 Software Conventions and Organization

In preparation for the polarimetric imaging efforts on M87, a leakage challenge was issued to confirm that all members are using the same conventions for the data, images and leakage estimates. We designed a simple synthetic data set, using very simplistic total intensity and polarization structure, and issued the challenge on July 1, 2019. In Figure 6.20, we show the true leakages and the convention test true image, composed of two small Gaussians of $5\,\mu as$ FWHM each, separated by $40\,\mu as$. The source on the right is $50\%$ polarized, with an EVPA of $45^\circ$. The only common aspect between the convention test and real EHT data is that we utilized the coverage for M87 observations on 2017 April 11. The D-term inputs to the convention test are not related to the EHT, but are chosen to remain within $15\%$, the expected range for EHT station D-terms. Station behavior is assumed to be perfect (no visibility amplitude gain variations), and the visibility phases have been self-calibrated to the true image to facilitate total intensity image reconstruction. The ground truth leakages were also provided to the participants for comparison.

Following the initial deadline for the convention test, on July 11, a comparison of D-term estimates from the individual participants showed some remaining scatter in various user-based choices (see Figure 6.21, left panel). Despite the scatter in D-term estimates, the reconstructed images were all consistent with the ground truth. Every member was able to recover the polarization and EVPA properties of the data set. The majority of submissions were using the LPCAL tool.
6.2 Polarimetric Imaging of M87

Figure 6.20: Ground truth for the convention test synthetic data set. *Left:* Truth image, composed of two Gaussians of 5 µas FWHM each, separated by 40 µas. The right Gaussian is 50% polarized, with an EVPA of 45°. *Right:* True input D-terms for all stations present in a typical EHT M87 coverage (no SPT).

Figure 6.21: Convention test leakage estimates for various individual submissions (colors) compared to the truth values in the synthetic data set (black). Symbols vary per station, filled symbols denote R polarization D-terms, hollow symbols denote L polarization D-terms. *Left:* Submissions to the convention test prior to the polarimetry workshop. *Right:* Submissions to the convention test on the first day of the polarimetry workshop, following a series of tutorials with the three leakage-estimation packages (*eht-imaging*, *polsolve*, *LPCAL*).
The same convention test was used as a simple training data set to perform tutorials on the various leakage-estimation software. Participants of the workshop were encouraged to submit leakage estimates using different software. In the right panel of Figure 6.21, we show the submissions after the tutorial sessions on the first day of the workshop, separated by software used. The distributions of D-terms have tightened, this is due to the use of similar scripts by the participants, based on procedures presented at the different tutorials. The SMA D-terms remain difficult to constrain, likely due to the short parallactic coverage and the missing SMA-JCMT cross-hand information. In Table 6.4, we show the mean D-terms derived per software and the true D-term values in the synthetic data set for comparison. We designed and implemented an online submission process and an automated comparison framework to ease the iterations of comparisons between software teams, both for synthetic data sets where the ground truth image and leakages are known and for the EHT M87, 3C 279, J1924–2914, and NRAO 530 data sets.

6.2.5 Leakage and Gain Calibration Strategy

In the imaging methods we divide the polarimetric calibration procedure for EHT data into two steps. In the first step, we 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 before polarimetric imaging and D-term calibration of the remaining stations. Baselines to the JCMT (which are redundant with SMA baselines) are removed from the data sets, to reduce complications from handling single-polarization data. The ALMA, APEX, and SMA D-terms are fixed in imaging with eht-imaging and polsolve; because LPCAL is unable to fix D-terms of specific stations to zero, it derives a residual leakage for these stations, which remains small.\(^{11}\) 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 intra-site baseline approach and instead solve for all D-terms (and station gains) starting with the base data product described in Section 6.2.2.

The point source assumption adopted in the imaging method intra-site baseline D-term calibration step is an extension to the intra-site redundancies already exploited in the EHT network calibration [Event Horizon Telescope Collaboration et al., 2019c], 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 \(\rho_{jk}\) in Equation 6.6, 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 \(\rho_{jk}\) 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, and NRAO 530 data sets).

\(^{11}\) The non-zero LPCAL D-terms for ALMA, APEX, and SMA indicate that there may either be possible residual leakage after intra-site baseline fitting or that uncertainties in the LPCAL estimates originate from e.g., a breakdown of the similarity approximation.
Table 6.4: Comparison of the mean and deviation D-terms from each software to the true values in the convention test synthetic data set. The full set of submissions from polarization effort members was used to derive these values.

<table>
<thead>
<tr>
<th>Site</th>
<th>True D-term</th>
<th>eht-imaging</th>
<th>polsolve</th>
<th>LPCAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA R</td>
<td>-0.0749 - 0.0479j</td>
<td>(-0.0796 - 0.0500j) ± (0.0039 + 0.0008j)</td>
<td>(-0.0771 - 0.0522j) ± (0.0031 + 0.0055j)</td>
<td>(-0.0704 - 0.0471j) ± (0.0012 + 0.0017j)</td>
</tr>
<tr>
<td>ALMA L</td>
<td>0.0679 - 0.0920j</td>
<td>(0.0700 - 0.0701j) ± (0.0016 + 0.0482j)</td>
<td>(0.0660 - 0.0954j) ± (0.0055 + 0.0089j)</td>
<td>(0.0682 - 0.0918j) ± (0.0018 + 0.0032j)</td>
</tr>
<tr>
<td>APEX R</td>
<td>-0.0809 + 0.0364j</td>
<td>(-0.0777 + 0.0379j) ± (0.0027 + 0.0013j)</td>
<td>(-0.0827 + 0.0332j) ± (0.0079 + 0.0037j)</td>
<td>(-0.0815 + 0.0366j) ± (0.0031 + 0.0018j)</td>
</tr>
<tr>
<td>APEX L</td>
<td>0.0234 - 0.0136j</td>
<td>(0.0278 - 0.0092j) ± (0.0016 + 0.0031j)</td>
<td>(0.0257 - 0.0152j) ± (0.0041 + 0.0028j)</td>
<td>(0.0262 - 0.0152j) ± (0.0021 + 0.0014j)</td>
</tr>
<tr>
<td>SMT R</td>
<td>0.0160 + 0.0361j</td>
<td>(0.0132 + 0.0258j) ± (0.0044 + 0.006j)</td>
<td>(0.0211 + 0.0346j) ± (0.0043 + 0.0016j)</td>
<td>(0.0160 + 0.0273j) ± (0.0043 + 0.0036j)</td>
</tr>
<tr>
<td>SMT L</td>
<td>-0.0284 + 0.0103j</td>
<td>(-0.0434 + 0.0111j) ± (0.0093 + 0.005j)</td>
<td>(-0.0267 + 0.0082j) ± (0.0038 + 0.0047j)</td>
<td>(-0.0247 + 0.0081j) ± (0.0041 + 0.0016j)</td>
</tr>
<tr>
<td>LMT R</td>
<td>0.0897 + 0.0136j</td>
<td>(0.1063 + 0.0211j) ± (0.0105 + 0.006j)</td>
<td>(0.1033 + 0.0033j) ± (0.004 + 0.005j)</td>
<td>(0.0899 + 0.0137j) ± (0.0058 + 0.0016j)</td>
</tr>
<tr>
<td>LMT L</td>
<td>-0.0751 - 0.0159j</td>
<td>(-0.0756 - 0.0033j) ± (0.0069 + 0.0072j)</td>
<td>(-0.0743 - 0.0211j) ± (0.0039 + 0.003j)</td>
<td>(-0.0668 - 0.0138j) ± (0.0057 + 0.0031j)</td>
</tr>
<tr>
<td>PV R</td>
<td>0.0540 - 0.0217j</td>
<td>(0.0473 - 0.0118j) ± (0.0077 + 0.0053j)</td>
<td>(0.0452 - 0.0234j) ± (0.0066 + 0.0032j)</td>
<td>(0.0465 - 0.0191j) ± (0.0016 + 0.0016j)</td>
</tr>
<tr>
<td>PV L</td>
<td>-0.0153 - 0.0116j</td>
<td>(-0.0204 - 0.0103j) ± (0.0057 + 0.002j)</td>
<td>(-0.0091 - 0.0129j) ± (0.0044 + 0.0024j)</td>
<td>(-0.0093 - 0.0136j) ± (0.0050 + 0.0033j)</td>
</tr>
<tr>
<td>SMA R</td>
<td>0.1278 + 0.0957j</td>
<td>(0.1158 + 0.0855j) ± (0.0112 + 0.0071j)</td>
<td>(0.1362 + 0.1184j) ± (0.0072 + 0.0132j)</td>
<td>(0.1283 + 0.0979j) ± (0.0027 + 0.0063j)</td>
</tr>
<tr>
<td>SMA L</td>
<td>0.0271 + 0.0632j</td>
<td>(0.0215 + 0.0568j) ± (0.0068 + 0.0086j)</td>
<td>(0.0246 + 0.0684j) ± (0.0057 + 0.0073j)</td>
<td>(0.0219 + 0.0630j) ± (0.0077 + 0.0007j)</td>
</tr>
</tbody>
</table>
Figure 6.22: Left panel: D-term estimates for ALMA, APEX, JCMT, and SMA from \texttt{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, but only one polarization is shown for JCMT and SMA per day
due to JCMT polarization setup limitations. Station averages across days and high/low
bands are shown as solid points with error bars. The depicted D-terms are provided in	tabulated form in Appendix D of Paper VII. Right panels: Fiducial D-terms for LMT,
PV, and SMT derived from the low band data via leakage calibration in tandem with
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 poste-
rior exploration methods have errorbars corresponding to the $1\sigma$ standard deviations
estimated from the posterior distributions of the resulting D-terms.

J1924–2914, NRAO530, 3C273, 1055+018, OJ287 and CenA as shown in Appendix D of Paper VII) on each day in one single fit per band. The results of these fits per station, polarization, day, and band are presented in Figure 6.22 (left panel), where we also plot the mean and standard
deviation of the D-terms across all days and both bands for each station and polarization.

In addition to intra-site 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 fiducial
Stokes $I$ images of chosen calibrator sources. Given the extreme resolving power of the EHT
array, all available calibrators are resolved on long baselines. We must therefore select sources
that are best imaged by the EHT array; these are compact non-variable sources with sufficient
$(u,v)$ coverage. Four targets in the EHT 2017 observations fit these criteria: M87, 3C279, J1924–
2914 and NRAO530. Stokes $I$ images of M87 and 3C279 have been published (Event Horizon
Telescope Collaboration et al. [2019a]; Kim et al. [2020]). Final Stokes $I$ images for the Sgr A*
calibrators NRAO530 and J1924–2914 will be presented in upcoming publications (S. Issaoun et
al. \textit{in prep.}, S. Jorstad et al. \textit{in prep.}) but the best available preliminary images are used to
For M87, although multiple imaging packages and pipelines were utilized in the Stokes $I$ imaging process, the resulting final ‘fiducial’ images from each method are highly consistent at the EHT instrumental resolution [e.g., Event Horizon Telescope Collaboration et al., 2019d, Figure 15]. We therefore selected a set of Stokes $I$ 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 ($\sim 10 \, \mu\text{as}$), 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, thereby accounting for residual station gain variations in the data that make imaging challenging. Using these self-calibrated data sets allows the imaging methods to focus on accurate reconstructions of the polarimetric Stokes $Q$ and $U$ brightness distributions and D-term estimation.

Preliminary D-terms estimated by the three imaging methods before testing and optimizing imaging parameters on synthetic data are reported in Section 6.2.6. The right panels of Figure 6.22 show the final D-terms for LMT, PV, and SMT derived from the imaging and posterior modeling methods after optimization on synthetic data (see Section 6.2.7). To quantify the agreement (or distance in the complex plane) between D-term estimates from different methods we calculate $L_1$ norms. The $L_1$ norms averaged over left and right (also real and imaginary) D-term components, over all stations and over the four observing days, are all less than 1% for each pair of imaging methods (see Appendix E of Paper VII). The mean values of the D-terms from the posterior exploration methods correlate well with the D-terms estimated by the imaging methods. For each combination of imaging and posterior exploration method the station averaged $L_1$ norms range from 1.5% to 1.89%. For verification purposes, we also estimate D-terms using data of several calibrator sources. We find that the D-terms derived by polarimetric imaging of these other sources are consistent with those of M87 (Appendix 6.C). Finally, we note our estimated SMT D-terms are similar to those computed previously using early EHT observations of Sgr A* [Johnson et al., 2015].

### 6.2.6 Preliminary Polarimetric Results

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. These results are not blind tests in analogy to the initial stage of total intensity imaging (see Paper IV). Nonetheless, in this early stage of manual 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.

In Figure 6.23, we present our recovered total intensity and preliminary polarimetric images of M87 on April 11 produced by the three methods available when preliminary image reconstructions were conducted. In Figure 6.23, we also show the D-terms associated with these images. Each method reproduces consistent D-terms for all three 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 in Figure 6.23 revealed the main structure of the linearly polarized source and suggested consistency between different imaging methods. They strongly motivate the need for conducting full parameter surveys for each method to optimize the chosen imaging parameters and validate the results on synthetic data.

### 6.2.7 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 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). In VLBI imaging, these parameters are often simply set by the user given their experience on similar data sets, or based on what appears to produce an image that is a good fit to the data and 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 method and attempt to chose an optimal set of parameters based on their performance in recovering the source structure and input D-terms from several synthetic data models. Appendix G of Paper VII provides more detail on the individual parameter surveys performed by each method. 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 images generated using general relativistic magnetohydrodynamic (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 synthetic source models blurred to the EHT nominal resolution are displayed in the first column of Figure 6.24.

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 Section 6.2.6. Random residual D-terms for ALMA, APEX, JCMT, and SMT (reflecting possible errors in the intra-site...
Figure 6.23: Left: Preliminary total intensity images reconstructed with eht-imaging, polsolve, and LPCAL on April 11 low band data. 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/µas². Cuts were made to omit all regions in the images where Stokes 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.
Figure 6.24: 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 6.25: 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 6.24). 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 $L_1 \equiv |D - D_{\text{Truth}}|$ 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.

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 $I$ image reconstructed via the SMILI fiducial script for M87 developed in Paper IV.

In Figure 6.24, 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 $I$ and the ground truth $I_0$ images,

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

(6.8)

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 $p \equiv Q + iU$,

$$\langle \vec{P} \cdot \vec{P}_0 \rangle = \frac{\text{Re} \langle p p_0^* \rangle}{\sqrt{\langle p^2 \rangle} \sqrt{\langle p_0^2 \rangle}}.$$  

(6.9)

The real part is chosen to measure the degree of alignment of the polarization vectors $(Q, U)$. In both cases, images are first shifted to give the maximum correlation coefficient for Stokes $I$. Because Stokes $I$ image reconstructions are tightly constrained by an a priori known total image flux, the Stokes $I$ 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 $(\text{field of view})^{-1}$ 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. In Figure 6.25 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 and the ground truth values $D_{\text{Truth}}$ in each approach,
we calculate the $L_1 \equiv |D_i - D_{\text{Truth}}|$ 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.

### 6.2.8 Fiducial Polarimetric Images of M87

In Figure 6.26, we present the fiducial M87 linear polarimetric images produced by each method from the low band data on all four observing days. The fiducial images from each method are broadly consistent with those from the preliminary imaging stage shown in Figure 6.23 of Section 6.2.6.

Unless otherwise explicitly indicated, we display low-band results in the main text. The high-band results are given in Appendix I of Paper VII. We decided to keep the analysis of the high and low band data separate for several reasons. First, the main limitations in the dynamic range and image fidelity in EHT reconstructions arise from the sparse sampling of spatial frequencies, not the data S/N. Increasing the S/N by performing band averaging does not improve the dynamic range of the reconstructed images. Second, treating each band separately minimizes any potential chromatic effects that might add extra limitations to the dynamic range, such as intra-field differential Faraday rotation. Finally, separating the bands in the analysis allowed us to use the high band results as a consistency check on the calibration of the instrumental polarization and image reconstruction for the low band data. We perform this comparison of the results obtained at both bands in Appendix I of Paper VII. We conclude that both the recovered D-terms and main image structures are broadly consistent between low and high band.

The different 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 6.26, 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 $I$ image with the April 11 “consensus” image presented in Figure 15 of Paper IV. The blurring kernel FWHMs selected by this method are $19\mu\text{as}$ for eht-imaging, DMC, and THEMIS, $20\mu\text{as}$ for LPCAL, and $23\mu\text{as}$ for polsolve.

The M87 emission ring is polarized only in its south–west region and the peak fractional polarization at $\approx 20\mu\text{as}$ 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 across time, frequency, and imaging method. The
6.2 Polarimetric Imaging of M87

Figure 6.26: Fiducial polarimetric M87 images produced by five independent methods. The results from all imaging and posterior exploration pipelines are shown on the four M87 observation days for low band (the low and high band results are consistent, see Appendix I of Paper VII). Total intensity is shown in grayscale, polarization ticks indicate the EVPA, the tick length indicates linear polarization intensity magnitude (where a tick length of 10 \( \mu \)as corresponds to \( \sim 30 \mu \text{Jy}/\mu\text{as}^2 \) of polarized flux density), 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 the linear polarized intensity. The solid, dashed, and dotted contour levels correspond to linearly polarized intensity of 20, 10, and 5 \( \mu \text{Jy}/\mu\text{as}^2 \). Cuts were made to omit all regions in the images where Stokes \( I < 10\% \) of the peak brightness and \( p < 20\% \) of the peak polarized brightness. The images are all displayed with a field of view of 120 \( \mu \)as, 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 \( I \) image with the consensus Stokes \( I \) image of Paper IV.
Chapter 6: Imaging a Black Hole with the EHT

Figure 6.27: Fiducial M87 average images produced by averaging results from our five reconstruction methods (see Figure 6.26). Method-average images for all four M87 observation days are shown, from left to right. These images show the low band results; for a comparison between these images and the high band results, see Figure 28 in Appendix I of Paper VII. 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 $\mu$as. Top: Total intensity, polarization fraction, and EVPA are plotted in the same manner as in Figure 6.26. 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 $I$ results [Paper I; Paper IV].

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 southern part of the ring appears less polarized on the later days. This evolution in the polarized brightness is consistent with the evolution in the Stokes $I$ image apparent in the underlying closure phase data ([alias?], Figure 14; Paper IV, Figure 23). However, as with the Stokes $I$ 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 6.27, 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 intensity distribution. We adopt the images in Figure 6.27 as a conservative representation of our final M87 polarimetric imaging results. Discussions of image properties are provided in
Sections 5 and 6 of Paper VII.

6.A Multi-Day Gain Comparisons with 3C 279

In Section 6.1.4, we compared the residual amplitude gains for the LMT and SMT on April 5 derived from separately imaging M87 and 3C279. We now expand this comparison to all four days on which these sources were observed.

Figure 6.28 compares the interleaved multiplicative station gains for M87 and 3C279 for each day; Table 6.5 presents the gain statistics for the two sources and compares them to the expected amplitude calibration error budget [Event Horizon Telescope Collaboration et al., 2019c]. The trends of inferred residual gain amplitudes are consistent among the imaging pipelines, and those of M87 are similar to those of 3C279. The inferred gains for M87 at the LMT have large excursions not seen in the 3C 279 gain trends; these excursions are likely due to poor pointing and tracking. These large LMT correction factors (up to $|g_{\text{LMT}}|^{-1} \approx 6$) are also consistent with those estimated using the SMT–PV and LMT–PV crossing tracks.

6.B 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.

6.B.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., 2016b; Matthews et al., 2018]. The post-correlation conversion of the ALMA data from a linear basis into a circular basis has 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, $\Delta$, 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 (the frequency band used in the EHT observations) are rotated by 45 degrees with respect to their projection on the focal plane. This 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 $RL$ ($LR$) 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
Figure 6.28: Derived residual gains for the SMT (left) and LMT (right) using self-calibration to images of M87 (red) and 3C279 (blue). The M87 images are the fiducial images from each pipeline; the 3C279 images were reconstructed separately, using adapted imaging scripts. The particularly large excursions on the LMT M87 gains are often due to poor pointing. For instance, excursions at ≈6 UTC are from difficulties tracking the source during transit (~81° elevation). Note the different ranges shown for each observing day.
### Table 6.5

<table>
<thead>
<tr>
<th>Station</th>
<th>April 5</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>A priori budget (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA</td>
<td>0.97±0.02</td>
<td>0.97±0.02</td>
<td>0.98±0.01</td>
<td>0.99±0.02</td>
<td>1.00±0.04</td>
<td>0.97±0.03</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>APEX</td>
<td>1.05±0.02</td>
<td>1.02±0.01</td>
<td>1.01±0.01</td>
<td>0.99±0.00</td>
<td>0.90±0.00</td>
<td>1.05±0.02</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>1.13±0.02</td>
<td>1.02±0.01</td>
<td>0.99±0.02</td>
<td>1.06±0.02</td>
<td>0.97±0.02</td>
<td>1.04±0.02</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>JCMT</td>
<td>1.00±0.02</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.01</td>
<td>1.02±0.02</td>
<td>1.00±0.00</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>LMT</td>
<td>1.46±0.01</td>
<td>1.55±0.03</td>
<td>1.47±0.01</td>
<td>1.08±0.04</td>
<td>1.21±0.13</td>
<td>1.35±0.27</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>SMA</td>
<td>0.90±0.01</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.01±0.01</td>
<td>1.04±0.00</td>
<td>0.99±0.01</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>1.14±0.04</td>
<td>0.96±0.07</td>
<td>0.98±0.02</td>
<td>1.33±0.12</td>
<td>1.14±0.04</td>
<td>0.94±0.05</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>7.5</td>
</tr>
</tbody>
</table>

### April 6

<table>
<thead>
<tr>
<th>Station</th>
<th>April 6</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>A priori budget (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA</td>
<td>1.02±0.03</td>
<td>0.98±0.02</td>
<td>0.99±0.02</td>
<td>1.00±0.02</td>
<td>1.04±0.09</td>
<td>0.99±0.01</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>APEX</td>
<td>0.98±0.02</td>
<td>1.01±0.02</td>
<td>1.01±0.02</td>
<td>1.00±0.01</td>
<td>0.97±0.09</td>
<td>1.01±0.01</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>1.06±0.04</td>
<td>1.05±0.02</td>
<td>1.01±0.03</td>
<td>1.06±0.00</td>
<td>0.99±0.05</td>
<td>1.02±0.01</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>JCMT</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.01</td>
<td>1.00±0.00</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>LMT</td>
<td>2.06±0.33</td>
<td>2.30±0.16</td>
<td>2.16±0.44</td>
<td>1.34±0.30</td>
<td>2.13±0.37</td>
<td>1.33±0.20</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>SMA</td>
<td>1.01±0.01</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.01</td>
<td>1.04±0.00</td>
<td>0.99±0.00</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>0.98±0.03</td>
<td>0.97±0.01</td>
<td>0.99±0.06</td>
<td>0.81±0.12</td>
<td>1.05±0.01</td>
<td>0.91±0.06</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>7.5</td>
</tr>
</tbody>
</table>

### April 10

<table>
<thead>
<tr>
<th>Station</th>
<th>April 10</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>A priori budget (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA</td>
<td>1.03±0.02</td>
<td>0.99±0.00</td>
<td>0.99±0.01</td>
<td>1.00±0.03</td>
<td>1.08±0.06</td>
<td>0.99±0.01</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>APEX</td>
<td>0.99±0.02</td>
<td>1.02±0.01</td>
<td>1.01±0.01</td>
<td>1.00±0.00</td>
<td>0.94±0.06</td>
<td>1.00±0.02</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>1.03±0.01</td>
<td>1.03±0.00</td>
<td>0.99±0.02</td>
<td>1.03±0.02</td>
<td>0.97±0.04</td>
<td>1.00±0.03</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>JCMT</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.00</td>
<td>1.00±0.04</td>
<td>0.98±0.00</td>
<td>1.00±0.00</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>LMT</td>
<td>1.37±0.33</td>
<td>1.50±0.13</td>
<td>1.35±0.12</td>
<td>0.96±0.04</td>
<td>1.17±0.09</td>
<td>1.19±0.07</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>SMA</td>
<td>1.01±0.04</td>
<td>1.00±0.02</td>
<td>1.00±0.02</td>
<td>1.00±0.01</td>
<td>1.01±0.02</td>
<td>1.00±0.01</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>1.02±0.03</td>
<td>0.94±0.03</td>
<td>0.94±0.01</td>
<td>0.97±0.02</td>
<td>1.10±0.04</td>
<td>0.96±0.01</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>7.5</td>
</tr>
</tbody>
</table>

### April 11

<table>
<thead>
<tr>
<th>Station</th>
<th>April 11</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>DIFMAP</th>
<th>eht-imaging</th>
<th>SMILI</th>
<th>A priori budget (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALMA</td>
<td>1.00±0.02</td>
<td>1.00±0.01</td>
<td>1.00±0.01</td>
<td>1.01±0.01</td>
<td>1.12±0.03</td>
<td>0.99±0.01</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>APEX</td>
<td>1.00±0.01</td>
<td>1.00±0.01</td>
<td>1.00±0.01</td>
<td>1.00±0.01</td>
<td>0.90±0.01</td>
<td>1.01±0.01</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>0.92±0.04</td>
<td>1.02±0.03</td>
<td>0.95±0.01</td>
<td>1.01±0.03</td>
<td>1.01±0.03</td>
<td>1.03±0.03</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>JCMT</td>
<td>1.00±0.01</td>
<td>1.00±0.01</td>
<td>1.01±0.01</td>
<td>1.00±0.00</td>
<td>1.00±0.01</td>
<td>1.00±0.00</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>LMT</td>
<td>1.35±0.16</td>
<td>1.35±0.10</td>
<td>1.23±0.11</td>
<td>1.16±0.20</td>
<td>1.25±0.25</td>
<td>1.47±0.14</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>SMA</td>
<td>1.00±0.04</td>
<td>0.99±0.01</td>
<td>1.00±0.01</td>
<td>1.00±0.00</td>
<td>1.04±0.01</td>
<td>1.00±0.01</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>PV</td>
<td>1.00±0.05</td>
<td>0.92±0.02</td>
<td>0.95±0.02</td>
<td>0.91±0.04</td>
<td>1.09±0.12</td>
<td>1.00±0.02</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.90±0.03</td>
<td>1.00±0.05</td>
<td>7.5</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.5:** Residual gain corrections derived for M87 and 3C279 for observations on all days (2017 April 5, 6, 10, 11). Numbers indicate the median of $|g|$ with stated uncertainties corresponding to the 25th and 75th percentiles. The multiplicative gains on the visibilities for each station are computed by self-calibrating the network-calibrated data sets to the reconstructions after rescaling the intra-site baseline amplitudes to match the total flux of the reconstructions. The percentage deviation from unity can be compared to the expectations from the station-based a priori error budget on the visibility amplitudes derived from Chapter 5. The SPT is omitted from gain comparisons the first two days as it did not observe 3C279 on 2017 April 5, and it observed only a single 3C279 scan on 2017 April 6.
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 of Paper VII) gives us absolute EVPA values instantaneously.

6.B.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 [Event Horizon Telescope Collaboration et al., 2019c]. 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 sampling as for the rest of the EHT [Event Horizon Telescope Collaboration et al., 2019b], 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 Event Horizon Telescope Collaboration et al. [2019a,b,c,d,e,f] are based on 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.

6.B.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 6.6) 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 Event Horizon Telescope Collaboration et al. [2019c]) 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 6.2.5 and in Appendix D of Paper VII are corrected by this rotation.
6.C LMT, SMT and PV D-terms using calibrator data: synthetic data tests, expected uncertainties and convergence with M87 results

Figure 6.29: 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, 3C279, J1924–2914 while April 7 is shown for NRAO530. Color scales indicates fractional polarization amplitude \(|\vec{m}|\) in the range from 0 to 2. Bottom panels: sources field rotation angle \(\phi\) for each station as a function of time. The figure is analogous to Figure 6.19 for M87 on April 5, 6, 10 and 11.

6.B.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).

6.C 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: 3C279, observed contemporaneously with M87; and J1924–2914 and NRAO 530, observed contemporaneously with our second EHT primary target, Sgr A*, in the second half of each observing day. 3C279 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 \((u,v)\) coverage.

In Figure 6.29, 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). For comparison, the field angle coverage for M87 on April 11 is also shown. 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 6.29, top row, we present the \(|\tilde{m}|\) 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 fractions are 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, interferometric-ALMA measurements taken contemporaneously to our EHT campaign found that 3C 279 may have non-negligible Stokes \(V\) [see Appendix E.2 in Goddi et al., 2021], which breaks the Stokes \(V = 0\) assumptions made in most of our calibration and imaging pipelines. 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 6.29) and have negligible Stokes \(V\) as measured by interferometric-ALMA [Goddi et al., 2021], 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 blazars 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 imaging with their current EHT coverage may not capture the complexity of the extended jets in these sources. Nevertheless, their weak polarization allows for better D-term estimates despite uncertainty in modeling their structure.

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 6.2.7) and add a seventh model constructed with ten Gaussian sources of varying total and polarization intensity, with some polarization structure offset from Stokes \(I\). This seventh data set is designed to mimic the basic structure seen in the preliminary polarimetric images of the two calibrators, for which the final images will be presented in forthcoming 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 of Paper VII.

In Figures 6.30 and 6.31 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 \(I\) and linear polarization \(P\)
6.C LMT, SMT and PV D-terms using calibrator data: synthetic data tests, expected uncertainties and convergence with M87 results

images are provided. Consistent with the results with M87 coverage, the Stokes $I$ 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 6.32 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 for the calibrators.

Finally, we estimate LMT, SMT and PV D-terms via polarimetric imaging of the J1924–2914 and NRAO 530 EHT data. 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 6.33, 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 the M87 and J1924–2914 D-terms compared to between M87 and NRAO 530. The calibrators have sparser ($u,v$) coverage (fewer scans), a narrower field rotation range, and more complex Stokes $I$ (extended structure and higher noise level) and polarimetric images compared to M87, which all impact the quality of our D-term estimation. 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) and that M87 itself is the best source for polarimetric leakage calibration.

Furthermore, while imaging calibrators we found that the quality of the Stokes $I$ image is critical for calibration. Both NRAO 530 and J1924–2914, as blazar sources, have complex jet structure that is not fully recovered with the current EHT coverage, and thus our Stokes $I$ 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. Assumptions about the Stokes $I$ image affect the results of the polarimetric imaging and calibration methods, for example in the self-similarity assumption employed for CLEAN reconstructions in our sub-component methods (see Appendix K of Paper VII).
Figure 6.30: 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.
Figure 6.31: 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 6.32: 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 NRAO530 on April 7 are shown in the top and bottom rows, respectively). Each data point represents the mean and standard deviation for each D-term estimate derived from synthetic data sets 1-7. The norm \( L_1 \equiv |D - D_{\text{Truth}}| \) is averaged over left, right, real, and imaginary components of the D-terms and over the four EHT stations shown.

Figure 6.33: Comparison of fiducial D-terms for the telescopes LMT, SMT, and PV estimated from M87 (April 11), J1924–2914 (April 11) and NRAO530 (April 7) low band data sets using the eht-imaging, polsolve, and GPCAL pipelines. 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/NRAO530 are averaged over different methods. LMT and SMT D-terms derived from J1924–2914 are found to be highly consistent with those from M87. The D-terms derived from NRAO530 imaging on average show larger deviation from M87 the D-terms; in particular, the PV D-terms estimated by eht-imaging show the largest deviation from all other estimates.
A CONCEPTUAL OVERVIEW OF SINGLE-DISH ABSOLUTE AMPLITUDE CALIBRATION


Event Horizon Telescope Memo Series, Memo 2017-CE-02.
released Sept 15 2017

Abstract

This document presents an outline of common single-dish calibration techniques and key differences between centimeter-wave and millimeter-wave observatories in naming schemes and measured quantities. It serves as a conceptual overview of the complete single-dish amplitude calibration procedure for the Event Horizon Telescope, using the Submillimeter Telescope (SMT) as the model station.

Note: This document is not meant to be used as a general telescope guide or manual from an engineering perspective. It contains a number of common approximations used at observatories as an attempt to reason through the methods used and the specific calibration information needed to calibrate VLBI amplitudes from Event Horizon Telescope observing runs. This document can be used in conjunction with similar calibration outlines from other stations for procedural comparisons.

Relevant terminology

Relevant variables introduced in this document (brightness temperatures approximated with the Rayleigh-Jeans approximation):

- $C_{\text{hot}}$: Counts measured when looking at the hot load (vane)
- $C_{\text{cold}}$: Counts measured when looking at the cold load (liquid nitrogen)
Appendix A: Amplitude Calibration Memo

- $C_{on}$: Counts measured observing a target
- $C_{sky}$: Counts measured when looking at blank sky
- $T_{cold}$: temperature of the cold load
- $T_{rx}$: receiver noise temperature
- $T_{amb}$: ambient temperature around the observatory, as measured by a weather station (physical temperature)
- $T_{sky}$: temperature of the atmospheric emission (the brightness temperature of the sky)
- $T_{cab}$: physical temperature of the receiver cabin (this is assumed to be the same as the ambient temperature)
- $T_{cal}$: derived temperature to give a correct temperature scale for the signal band
- $T_{inject}$: injected known temperature (of calibrator or noise diode) in the signal chain
- $T_{sys}$: system noise temperature of the system
- $T^{*}_{sys}$: effective system noise temperature (corrected for atmospheric attenuation)

Efficiency and correction terms:

- $r_{sb}$: sideband ratio - since the SMT has a sideband-separating receiver, $r_{sb} = \frac{2g_{i}}{g_{s}} \ll 1$ since no signal comes from the image band but some leakage can still be present
- AM: amount of airmass in the line of sight of the receiver (elevation-dependent)
- $\tau_{0}$: atmospheric opacity at the zenith
- $e^{-\tau}$: atmospheric attenuation factor, which damps the signal based on atmospheric opacity in the line of sight $\tau = \tau_{0} \times AM$
- el: elevation of the antenna dish for a particular observation (in degrees)
- $g(\text{el})$: elevation-dependent gain curve correcting for changing illumination of the main reflector and ground contributions as the dish moves and tilts to different elevations
- $\eta$: forward efficiency representing the fraction of power received through the forward atmosphere (accounting for rearward losses)
- $\eta_{taper}$: efficiency loss due to non-uniform illumination of the aperture plane by the tapered radiation pattern
- $\eta_{block}$: aperture blockage efficiency due to blocking of the feed by the sub-reflector (including its support legs)
- $\eta_{spillover}$: feed spillover efficiency past the main reflector — it is the ratio of the power intercepted by the reflective elements to the total power
A.1 Introduction to standard single-dish $T_{\text{sys}}$ calibration techniques

The following is an outline of the different calibration procedures for cm-wave and mm-wave observatories and the different quantities they output. The equations provided here contain various approximations commonly used but are not exact from an engineering perspective. They are only meant to serve as guidelines for a quick understanding of the outputs of the two different techniques.

A.1.1 The antenna-based system-equivalent flux densities (SEFDs)

A telescope’s system-equivalent flux density (SEFD) is simply the noise contribution of the system, given by the system noise temperature, and all losses and gains, converted to a flux density scale. The SEFDs can be calculated using system noise temperature $T_{\text{sys}}$ measurements and all efficiencies and contributions to source attenuation and noise, and one can determine the sensitivity of the telescope when compared to other telescopes in the array. The higher a telescope’s SEFD, the lower its sensitivity. Ultimately, the flux density of a source is simply the telescope’s SEFD, which contains all system and telescope parameters and efficiencies, multiplied by the ratio of signal to noise power (defined as $r_{S/N}$) of the source detection. The equation for the SEFD can be subdivided into three main components, each with station-based variations for how they are determined and measured. The three components to the SEFD are:

1. $T_{\text{sys}}$: the total noise characterization of the system, given by the system noise temperature

2. $\frac{e^r}{\eta_8}$: the correction terms for attenuation of the source signal by the atmosphere and rearward losses (ohmic losses, rearward spillover and scattering) of the telescope

3. $G$: The antenna gain, including all the loss terms from the telescope and the conversion from a temperature scale (K) to a flux density scale (Jansky), given by the “degrees per flux density unit" factor (DPFU) in K/Jy and the normalized elevation-dependent gain curve $g(\text{el})$: $G = \text{DPFU} \times g(\text{el})$

This gives the following general equation for a telescope’s SEFD [Kraus, 1966; Burke & Graham-Smith, 2009; Wilson et al., 2013; Thompson et al., 2017]:

$$\text{SEFD} = \frac{T_{\text{sys}} e^r}{\eta G} \quad \text{(A.1)}$$
The flux density of a source detected with a given ratio of signal to noise power \( r_{S/N} \) is then:

\[
S_{\text{source}} = \text{SEFD} \times r_{S/N} = \frac{r_{S/N} \times T_{\text{sys}} e^\tau}{\eta G} \tag{A.2}
\]

For mm-observatories, which measure the effective system noise temperature \( T^*_\text{sys} = T_{\text{sys}} e^\tau \) directly using the chopper technique (explained in the next section), the SEFD equation can be rewritten in only two components, the effective system noise temperature and the antenna gain:

\[
\text{SEFD} = \frac{T^*_\text{sys}}{G} \tag{A.3}
\]

For the SMT, the SEFD at zenith is of order 13 000 Jansky.

### A.1.2 The receiver noise temperature

#### Two-load (hot and cold) calibration

During a two-load calibration (also called cold calibration), the Y-factor and the receiver noise temperature are measured using voltage or counts measurements with a hot and a cold load. In principle the receiver noise temperature can be estimated from \( T_{\text{sys}} \) measurements at very low opacities (\( \tau \ll 1 \)) by extrapolating a linear fit of airmass versus \( T_{\text{sys}} \) to zero airmass. However, it is highly recommended to measure a receiver noise temperature at least once an observing night, as this yields more accurate \( T_{\text{sys}} \) measurements rather than backtracking in post-processing. The Y-factor is calculated with the following:

\[
Y = \frac{C_{\text{hot}}}{C_{\text{cold}}} \tag{A.4}
\]

where the numerator is \( C_{\text{hot}} \), the counts obtained from the hot load and the denominator is \( C_{\text{cold}} \), the counts obtained from the cold load. The Y-factor also enables an easy diagnostic of the sensitivity of the receiver. A high Y-factor means little receiver noise, and thus sensitive observations (of course what constitutes “high” depends on the type of receiver and the observing frequency).

Then the receiver noise temperature is determined as follows:

\[
T_{\text{rx}} = \frac{T_{\text{hot}} - YT_{\text{cold}}}{Y - 1} \tag{A.5}
\]

Here temperatures are used, where \( T_{\text{hot}} \) is the temperature of the hot load (for the SMT this is near room temperature \( \sim 290 \text{K} \)) and \( T_{\text{cold}} \) is the cold load temperature (for the SMT this is the temperature of liquid nitrogen \( \sim 77 \text{K} \)).

### A.1.3 The effective system noise temperature

A lot of confusion comes from mixtures of complicated calibration documents for different types of calibrations. This section is an attempt to approximately explain two of the common techniques for \( T_{\text{sys}} \) measurements (chopper wheel common for mm-telescopes, direct for cm-telescopes), what they output and what they mean for data processing. The following outlines are modelled for an SMT-like telescope, thus with a sideband-separating (SSB) receiver.
Clarification of system temperature jargon

We define the system noise temperature as the contributions by the receiver and the sky to a source measurement (assuming $T_{\text{CMB}}$ is negligible), where $\eta$ is the forward efficiency, accounting for rearward efficiency loss due to ohmic losses, rear spillover and scattering [e.g., Baars, 1973; Gordon et al., 1992; Greve & Bremer, 2010]:

$$T_{\text{sys}} = T_{\text{rx}} + T_{\text{sky}} = T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau}) \quad (A.6)$$

Before entering the atmosphere, the source signal is defined as $\text{Sig} = T_{\text{source}}$. After attenuation by the atmosphere, the signal becomes $\text{Sig} = \eta e^{-\tau} T_{\text{source}}$, where the exponential is the atmospheric attenuation factor ($\tau$ is the opacity in the line of sight) and $\eta$ embodies rearward efficiency losses. Therefore, the ratio of signal to noise power of a telescope must depend on this received signal (not taking into account ground and ambient contributions):

$$r_{S/N} = \frac{\eta e^{-\tau} T_{\text{source}}}{T_{\text{sys}}} = \frac{\eta T_{\text{source}}}{e^{\tau} T_{\text{sys}}} \quad (A.7)$$

We thus define the effective system noise temperature $T^*_{\text{sys}}$:

$$T^*_{\text{sys}} = \frac{e^{\tau}}{\eta} T_{\text{sys}} = T_{\text{rx}} \frac{e^{\tau}}{\eta} + T_{\text{atm}} \left(\frac{e^{\tau}}{\eta} - 1\right) \quad (A.8)$$

The effective system temperature is the best description of the sensitivity of a telescope: the system sensitivity drops rapidly (exponentially) as opacity increases.

Direct (switched noise diode) method

This method is commonly used at cm-observatories, such as the VLBA. The system noise temperature is obtained using a known source or a switched noise diode with a known temperature placed in the signal chain. The equation is the following, where $C_{\text{sky}}$ represents the counts on blank sky, so only receiver noise and sky contribute, and $C_{\text{on,cal}}$ represents counts on the calibrator (or diode), such that the signal contains the source, the receiver and the sky. $T_{\text{inject}}$ is the temperature of the diode or the brightness temperature of the source (known for common calibrators), which turns the counts scale to a temperature scale. When the telescope is pointed at blank sky in the calibration procedure, without the source signal, the temperature contribution is entirely noise from the receiver and the atmospheric emission, and thus is the system noise temperature $T_{\text{sys}}$:

$$T_{\text{off,cal}} = T_{\text{rx}} + T_{\text{sky}} = T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau}) \quad (A.9)$$

When the telescope is pointed at the calibrator (or diode) of a known brightness (or physical) temperature, the source signal is added to the temperature contribution:

$$T_{\text{on,cal}} = T_{\text{rx}} + T_{\text{sky}} + T_{\text{inject}} \quad (A.10)$$

The system temperature is then determined in the following way:

$$T_{\text{sys}} = \frac{C_{\text{sky}}}{C_{\text{on,cal}} - C_{\text{sky}}} T_{\text{inject}} = \frac{T_{\text{off,cal}}}{T_{\text{on,cal}} - T_{\text{off,cal}}} T_{\text{inject}} \quad (A.11)$$

$$= \frac{(T_{\text{rx}} + T_{\text{sky}})}{(T_{\text{rx}} + T_{\text{sky}} + T_{\text{inject}}) - (T_{\text{rx}} + T_{\text{sky}})} T_{\text{inject}} = \frac{(T_{\text{rx}} + T_{\text{sky}})}{T_{\text{inject}}} T_{\text{inject}} = T_{\text{rx}} + T_{\text{sky}}$$

217
Since the brightness temperature of the source observed (or the diode temperature) is determined outside the atmosphere, the system noise temperature calculated with this method does not include effects on sensitivity due to atmospheric attenuation \((e^\tau\text{ term})\). This is because the contribution of the source or diode is added to the signal chain (as opposed to the chopper technique that blocks everything but the receiver noise, explained and derived in the next section). This method does not provide an effective system temperature directly, only the receiver and sky contributions to the noise (which cannot be disentangled from each other).

In order to obtain the effective system temperature, opacity measurements during observations must be obtained. This is done either by using water vapor radiometers (or tipping radiometers) or by using the telescope as a tipper using sky tips. Tipping radiometers are notoriously unreliable (although water vapor radiometers perform very well), and sky tips must be done very often (every 10 min) and take up valuable observing time. This makes this method highly cumbersome for frequencies at which the atmosphere cannot be neglected.

The effective system temperature is thus:

\[
T_{\text{sys}}^* = e^\tau \frac{\eta}{T_{\text{sys}}},
\]

such that the effective (opacity-corrected) antenna temperature of a source (where \(C_{\text{on}}\) is the telescope signal on target) can be given using ON-OFF measurements as:

\[
T_A^* = \frac{C_{\text{on}} - C_{\text{sky}}}{C_{\text{sky}}} T_{\text{sys}}^* = \frac{T_{\text{on}} - T_{\text{off}}}{T_{\text{sys}}} T_{\text{sys}} = e^\tau \frac{\eta}{\tau} T_A
\]

### Chopper (or single load) calibration

The chopper (or single load) calibration technique is commonly used by (sub)mm observatories. The system noise temperature is obtained by placing an ambient temperature load \(T_{\text{hot}}\) that has properties similar to a blackbody in front of the receiver, blocking everything but the receiver noise. As long as \(T_{\text{atm}} \sim T_{\text{hot}}\), this method automatically compensates for rapid changes in mean atmospheric absorption [Berdahl & Fromberg, 1982; Berger et al., 1984; Berdahl & Martin, 1984].

For calibration of source measurements, we want to obtain the effective sensitivity of the system, not a comparison between the receiver and sky contributions to noise. Therefore, we want to obtain the effective system noise temperature \(T_{\text{sys}}^*\) to calibrate source measurements.

To first order, the chopper method directly measures \(T_{\text{sys}}^*\). This is obtained via the following equation:

\[
T_{\text{sys}}^* = T_{\text{hot}} \frac{C_{\text{sky}}}{C_{\text{hot}} - C_{\text{sky}}} = T_{\text{rx}} e^\tau \frac{\eta}{\tau} + T_{\text{atm}} (e^\tau - 1),
\]

where \(C_{\text{sky}}\) is the voltage/count signal on blank sky and \(\tau\) is the opacity in the line of sight.

**How does the chopper technique directly provide \(T_{\text{sys}}^*\)?** This is shown simply by investigating the exact output by the chopper technique. The chopper system temperature equation is given in telescope counts, where \(C_{\text{hot}}\) are the counts measured when the blocker/chopper/vane is in place, and \(C_{\text{sky}}\) is our usual blank sky counts. In terms of temperatures, the temperature contribution when the blocker is in place \(T_{\text{block}}\) is defined as:

\[
T_{\text{block}} = T_{\text{rx}} + T_{\text{hot}},
\]

218
where $T_{\text{hot}}$ is the temperature of the hot load itself. The load completely blocks the sky emission, which changes the calibration equations from the direct (or diode) calibration method. As seen in the direct method, the blank sky contribution is simply the system noise temperature:

$$T_{\text{off}} = T_{\text{rx}} + T_{\text{sky}} = T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau})$$  \hspace{1cm} (A.16)

We can thus write the chopper equation (eq. A.14) in terms of temperatures:

$$T_{\text{sys}}^* = T_{\text{hot}} \frac{C_{\text{sky}}}{C_{\text{hot}} - C_{\text{sky}}} = T_{\text{hot}} \frac{T_{\text{off}}}{T_{\text{block}} - T_{\text{off}}}$$ \hspace{1cm} (A.17)

$$= T_{\text{hot}} \frac{T_{\text{rx}} + T_{\text{sky}}}{(T_{\text{rx}} + T_{\text{hot}}) - (T_{\text{rx}} + T_{\text{sky}})}$$ \hspace{1cm} (A.18)

We assume the hot load is at ambient temperature, and so $T_{\text{hot}} = T_{\text{amb}} = T_{\text{atm}}$. This gives:

$$T_{\text{sys}}^* = T_{\text{atm}} \frac{T_{\text{rx}} + T_{\text{sky}}}{(T_{\text{rx}} + T_{\text{atm}}) - (T_{\text{rx}} + T_{\text{sky}})}$$ \hspace{1cm} (A.19)

As we have defined $T_{\text{sky}} = T_{\text{atm}}(1 - \eta e^{-\tau})$, we can simplify:

$$T_{\text{sys}}^* = T_{\text{atm}} \frac{T_{\text{rx}} + T_{\text{sky}}}{T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau})}$$ \hspace{1cm} (A.20)

$$= T_{\text{atm}} \frac{T_{\text{rx}} + T_{\text{atm}}}{T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau})}$$ \hspace{1cm} (A.21)

Finally we obtain:

$$T_{\text{sys}}^* = T_{\text{hot}} \frac{C_{\text{sky}}}{C_{\text{hot}} - C_{\text{sky}}} = T_{\text{rx}} + T_{\text{atm}}(1 - \eta e^{-\tau})$$ \hspace{1cm} (A.26)

$$= T_{\text{rx}} e^{\tau} + T_{\text{atm}} \left( \frac{e^{\tau}}{\eta} - 1 \right) = \frac{e^{\tau}}{\eta} (T_{\text{rx}} + T_{\text{sky}})$$ \hspace{1cm} (A.27)

If we compare the chopper effective system noise temperature to the system temperature from the direct method:

$$T_{\text{sys}}^* = \frac{T_{\text{sys}}}{\eta e^{-\tau}}$$ \hspace{1cm} (A.28)

To first order, the chopper calibration (or alternatively named the single-load calibration) corrects for atmospheric attenuation of an observed source and rearward losses of the telescope by directly measuring $T_{\text{sys}}^*$. It is also worth noting that during VLBI observing, the quarter wave-plate is added to the signal chain to convert linear to circular polarization: any losses associated with the addition of the wave-plate will be automatically calibrated and included in the $T_{\text{sys}}^*$ measurement from the chopper technique in the same way as the atmospheric and rearward losses.
A.1.4 Getting a flux density

We have defined the antenna temperature (modified for measured quantities at a telescope) as:

\[ T_A^* = \frac{C_{\text{on}} - C_{\text{sky}}}{C_{\text{sky}}} T_{\text{sys}} \]  

(A.29)

To get a flux density, we must correct for the aperture efficiency \( \eta_A \) (determined through different loss terms or planet flux measurements) and gain curve \( g(\text{el}) \) as a function of elevation of the telescope and convert from a temperature scale to a flux density scale (where \( k \) is the Boltzmann constant), dependent on the geometric area of the dish \( A_{\text{geom}} \):

\[ S = \frac{T_A^*}{\eta_A g(\text{el}) A_{\text{geom}}} \times \frac{2k}{T_{\text{sys}}} \]  

(A.30)

The equation above is then the final expression to obtain a flux density for a given source. If we expand using all the terms we’ve discussed, we get the following:

\[ S = \frac{T_A^*}{\eta_A g(\text{el}) A_{\text{geom}}} \times \frac{2k}{T_{\text{sys}}} \times \frac{C_{\text{on}} - C_{\text{sky}}}{C_{\text{sky}}} \times \frac{2k}{A_{\text{geom}}} \]  

(A.31)

\[ \eta_A g(\text{el}) \times \frac{2k}{T_{\text{sys}}} \times \frac{C_{\text{on}} - C_{\text{sky}}}{C_{\text{sky}}} \times \frac{2k}{A_{\text{geom}}} \]  

(A.32)

\[ \frac{T_{\text{on}} - T_{\text{off}}}{T_{\text{off}}} \times \frac{T_{\text{sys}}}{\eta_A g(\text{el}) \eta e^{-\tau} A_{\text{geom}}} \]  

(A.33)

Now the flux density is rewritten also in terms of a system noise temperature determined with the direct method.

We can subdivide the flux density equation into three major parts:

1. The ratio of signal to noise power of the observed source as measured by the telescope (thus attenuated by the atmosphere): \( T_{S/N} = \frac{T_{\text{on}} - T_{\text{off}}}{T_{\text{off}}} \)

2. The total noise characterization of the system, including the correction term for atmospheric absorption, given by the effective system noise temperature: \( T_{\text{sys}}^* = e^{\tau} T_{\text{sys}} \)

3. The antenna gain \( G \), including all the loss terms from the telescope and the conversion from a temperature scale (K) to a flux density scale (Jansky), given by the “degrees per flux density unit’ factor (DPFU) and the gain curve:

\[ \text{DPFU} = \frac{\eta_A A_{\text{geom}}}{2k} \text{ giving } G = \text{DPFU} \times g(\text{el}) \]

We can thus simplify the flux density equation using the three main terms actually measured by the SMT:

\[ S = \frac{r_{S/N} \times T_{\text{sys}} e^{\tau}}{\eta G} = \frac{r_{S/N} \times T_{\text{sys}}^*}{G} \]  

(A.34)

Determining a gain curve

As previously mentioned, the characterization of the antenna gain \( G \) is subdivided into two quantities that must be separately provided for the calculation of the SEFDs: the gain curve \( g(\text{el}) \)
A.1 Introduction to standard single-dish $T_{\text{sys}}$ calibration techniques

and the DPFU (explained in the next section). The characterization of the telescope’s geometric (opacity-free) gain curve is an important part of the flux density calibration, and is particularly crucial for the EHT a priori amplitude calibration due to the low-elevation observations of some science targets (including Sgr A*) for the northern hemisphere stations.

Telescopes do not have perfect surfaces, and must thus suffer some losses of signal due to distorted illumination of the main reflector as they slowly move to different elevations. This large-scale surface deformation affects the received signal and is not taken into account in the measurements leading to the efficiency and DPFU characterization. These losses can be determined by tracking sources through a wide range of elevations, and thus measure an elevation-dependent gain curve for the telescope, where the maximum ($g = 1$) is set where the telescope is expected to be most efficient. The source measurements used to obtain a gain curve must of course be calibrated for all other effects, including telescope efficiency (through the DPFU) and the atmospheric attenuation of the signal (through $T_{\text{sys}}^*$). At the SMT, this is done by observing two sources (usually K3-50 and W75N, a planetary nebula and a star-forming region, due to their similar up-time plots and wide range of elevation) contiguously, tracked as they increase and decrease in elevation from the tree-line to transit and vice-versa.

The gain curve is estimated by fitting a polynomial (usually second-order for standard radio-dishes). If more than one source is used, this is done once the flux density measurements are normalized around a plateau (to a relative gain scale). This normalized gain curve must be written in the form of a second order polynomial (in the standard VLBA format for ANTAB), where ‘el’ is the elevation in degrees:

$$g(\text{el}) = a_2(\text{el})^2 + a_1(\text{el}) + a_0$$  \hfill (A.35)

Each parameter must not be rounded to the uncertainties of the fit but instead many significant figures should be provided. Uncertainties for each parameter as outputted by the polynomial fit must also be provided, along with the full covariance matrix of the fit parameters. This will help determine an error estimate for the gain curve and propagate to the error estimation of the final SEFDs. Additionally, a plot of the relative gains (normalized fluxes) versus elevation and the fitted polynomial should be provided if possible, as shown in Fig. A.1.

**Determining the DPFU**

The degrees per flux density unit (or DPFU) is the characterization of the temperature to flux density scale of a telescope. The DPFU is used to calibrate the telescope measurements to a flux density scale and is obtained using known flux calibrators, particularly planets, or by bootstrapping near-in-time observations of non-planet sources from telescopes with well-defined and accurate flux density measurements. This enables to check the flux density scale obtained by the telescope by directly measuring an aperture efficiency.

The DPFU is estimated with the following equation, where $k = 1.38 \times 10^{-23} \text{J/K} = 1.38 \times 10^3 \text{Jy/K}$:

$$\text{DPFU} = \frac{\eta A_{\text{geom}}}{2k} \left[ \text{K/Jy} \right]$$  \hfill (A.36)
The geometric area $A_{\text{geom}}$ is simply the area of the dish, where $D$ is the dish diameter:

$$A_{\text{geom}} = \frac{\pi D^2}{4}$$  \hspace{1cm} (A.37)

The aperture efficiency is the most difficult part of the estimation of the DPFU. It represents the efficiency of the telescope compared to a telescope with a perfect collecting area (uniform illumination, no blockage or surface errors) and it is determined using observations of known calibrator sources, usually planets. The observed planet fluxes are then compared to expected planet brightness temperatures from a planet simulation software for a perfect telescope at the given frequency and beam width.

The aperture efficiency $\eta_A$ is found using the following equation, where $T_A^*$ is the observed effective antenna temperature, $g(\text{el})$ is the telescope gain curve, $k$ is the Boltzmann constant, $A_{\text{geom}}$ is the geometric area of the telescope and $S_{\text{beam, sim}}$ is the expected flux density of the planet in the telescope beam from the simulation program used:

$$\eta_A = \frac{2k}{A_{\text{geom}} g(\text{el})} \frac{T_A^*}{S_{\text{beam, sim}}}$$  \hspace{1cm} (A.38)

Or similarly the DPFU is directly given by:

$$\text{DPFU} = \frac{T_A^*}{g(\text{el}) S_{\text{beam, sim}}}$$  \hspace{1cm} (A.39)

For extended sources, it is important to calibrate the flux density observed in the beam because some emission might not be picked up by the telescope. The aperture efficiency is only concerned
by the main beam flux density, and so the following equation is used to calibrate the simulated flux density in the beam for an extended source, where $S_{\text{sim}}$ is the expected total flux density of the source:

$$S_{\text{beam, sim}} = S_{\text{sim}} \times K$$  \hspace{1cm} (A.40)

Here $K$ is the following, where $\theta_{\text{mb}}$ is the half-power beam-width in arcseconds of the primary lobe of the telescope beam pattern (telescope beam diameter) and $\theta_s$ is the diameter in arcseconds of the observed extended source, usually given by the simulation program:

$$x = \frac{\theta_s}{\theta_{\text{mb}}} \sqrt{\ln(2)}$$  \hspace{1cm} (A.41)

$$K = 1 - e^{-x^2}$$  \hspace{1cm} (A.42)

This $K$ factor is the ratio of the beam-weighted source solid angle and the solid angle of the source on the sky. It is in fact the integral of the antenna pattern of the telescope (approximated as a normalized gaussian) $P(\theta, \phi) = e^{-\ln2(2\theta/\theta_{\text{mb}})^2}$ and a disklike source with a uniform brightness distribution $\Psi(\theta, \phi) = 1$ over the size of the extended source. This serves very well for our a priori calibration purposes\(^1\).

$$K = \frac{\Omega_{\text{sum}}}{\Omega_s} = \frac{1}{\Omega_s} \int_{\text{source}} P(\theta - \theta', \phi - \phi') \Psi(\theta', \phi') d\Omega'$$  \hspace{1cm} (A.43)

$$K = \frac{1}{\Omega_s} \int_{\text{source}} P(\theta - \theta', \phi - \phi') d\Omega'$$  \hspace{1cm} (A.44)

To minimize the number of approximations used by different planet simulation softwares, the expected total flux density can be estimated by:

$$S_{\text{sim}} = \frac{2h}{c^2} \frac{\nu^3 \Omega_s}{e^{\frac{h}{kT_B}} - 1},$$  \hspace{1cm} (A.45)

where $\nu$ is the observing frequency in Hz, $h$ is the Planck constant, $c$ is the speed of light (in m/s), $T_B$ is the mean brightness temperature for the planet (assuming a disk of uniform temperature) from the simulation program, and $\Omega_s$ is the solid angle of the source on the sky in steradians. Since we are dealing with very small objects, the latter can be approximated using the small angle approximation, where $\theta_s$ is the apparent diameter in radians of the planet observed:

$$\Omega_s \approx \frac{\pi \theta_s^2}{4}$$  \hspace{1cm} (A.46)

Of course this process heavily depends on assumptions made in the planet calibration, such as accurate predicted planet brightness temperatures from available software, telescope beam width used, stable weather conditions and a well-calibrated instrument in terms of pointing and focus [Bensch et al., 2001].

\(^1\)More detail on this method in *Calibration of spectral line data at the IRAM 30m radio telescope* by C. Kramer.
An average value for the aperture efficiency can be estimated from the individual measurements during a particular observing run, but it is preferable to keep the time-dependence of the variable if a telescope’s efficiency is expected to vary with temperature and sunlight, causing systematic differences in the telescope performance between day-time and night-time observing.

Even more preferable, a plot of long-term trends of the aperture efficiency, using additional measurements outside EHT observing or even from previous years, would greatly help understand the time-dependent nature of the aperture efficiency of a particular telescope. As the scatter between individual measurements can be caused by various factors, such as unstable weather or changing pointing/focus accuracy, it is not always representative of the true aperture efficiency change in the observations. A trend exhibited in the long-term as a function of time would be more reliable to estimate an aperture efficiency for a particular scan. Such a plot is shown in Fig. A.2, as an example from the JCMT.

If a UT time-dependence is found for a particular station, a fit for this dependence must be provided, as well as the covariance matrix for the fit parameters, for error analysis of the a priori deliverables. A fit to the UT time-dependence would be the most robust against various observing effects from day to day and session to session and should be very stable over the years, provided no major work has been done on the telescope. For telescopes with no visible time-dependence, a mean aperture efficiency (or DPFU) will suffice, with the appropriate error estimate.
We can also write the aperture efficiency as the combination of various individual forward efficiencies, each closely approximated for the telescope via various measurements:

\[ \eta_A = \eta_{\text{taper}} \times \eta_{\text{block}} \times \eta_{\text{spillover}} \times \eta_{\text{Ruze}} \]  

(A.47)

Each efficiency term corresponds to an aspect of the telescope feed:\(^2\):

- \( \eta_{\text{taper}} \) is the efficiency loss due to non-uniform illumination of the aperture plane by the tapered radiation pattern/feed function (also formally known as the illumination efficiency). It is the most important contributor to the aperture efficiency.

- \( \eta_{\text{block}} \) is the aperture blockage efficiency due to blocking of the feed by the sub-reflector (including its support legs)

- \( \eta_{\text{spillover}} \) is the feed spillover efficiency past the main reflector - it is the ratio of the power intercepted by the reflective elements beyond the edge of the sub-reflector and primary to the total power. It is due partly to cold sky and partly to a warm background, and is elevation-dependent.

- \( \eta_{\text{Ruze}} \) is the surface error efficiency (also called “Ruze loss” or scattering efficiency) calculated from Ruze’s formula [Ruze, 1952]. It is due to small scale, randomly distributed deviations of the reflector from the perfect paraboloidal shape. Ruze’s formula is presented below, where \( \sigma \) is the surface rms (accounting for small-scale deviations from a perfect surface through dish holography) and \( \lambda \) is the observing wavelength:

\[ \eta_{\text{Ruze}}(\lambda) = e^{-16\pi^2\sigma^2/\lambda^2} \]  

(A.48)

In summary, the aperture efficiency accounts for all forward losses of the telescope, which come from different contributions. As previously mentioned in section A.1.3, the chopper technique itself account for the rearward losses of the telescope automatically. These losses are also outlined in the following section.

**Other efficiencies**

The main beam efficiency of a telescope is the fraction of observed power in the main lobe of the telescope beam pattern. Let the beam solid angle (the full antenna pattern) be \( \Omega_A \) and the main beam solid angle (the main lobe) be \( \Omega_{\text{mb}} \). The main beam efficiency is written as the ratio between the total beam and main beam solid angles:

\[ \eta_{\text{mb}} = \frac{\Omega_{\text{mb}}}{\Omega_A} \]  

(A.49)

It is estimated with the following, where \( T_{\text{mb}} \) is the main beam temperature of a source that fills the main beam, as estimated from the simulation program:

\[ \eta_{\text{mb}} = \frac{S_{\text{beam, sim}} \eta_A A_{\text{geom}}}{T_{\text{mb}} \frac{2k}{T_{\text{mb}}}} \]  

(A.50)

It should be noted that the main beam efficiency is not the same as the aperture efficiency and should not be used to determine telescope DPFUs and SEFDs.

The forward efficiency $\eta_l$ represents the fraction of power received through the forward atmosphere (in other terms it is the coupling of the receiver to the cold sky) and is written as the ratio between the solid angle over the forward hemisphere of the telescope and the beam solid angle and it is typically close to unity (but drops with frequency due to loss of receiver sensitivity):

$$\eta_l = \frac{\Omega_{2\pi}}{\Omega_A}$$  \hspace{1cm} (A.51)

The only way to estimate it is via sky-dips, by measuring the atmospheric emission with elevation:

$$T_A(\text{el}) = \eta_l T_{\text{atm}}(1 - e^{-\tau/\sin(\text{El})}) + (1 - \eta_l) T_{\text{amb}}$$  \hspace{1cm} (A.52)

It is important to note that sky-dips measure both the atmospheric opacity and the forward efficiency so they need to be disentangled. Fortunately, this is not an issue for the EHT because the chopper technique implicitly corrects for the forward efficiency $\eta_l$ (see Section A.1.3).

### A.2 Miscellaneous explanations

#### A.2.1 VLBI and the Event Horizon Telescope array

**Determining the antenna-based SEFD for VLBI**

The SEFD needed for calibration of single-dish on-off observations and that for VLBI are identical. The equation for the antenna-based SEFD for VLBI observations is thus:

$$\text{SEFD} = \frac{T^*_{\text{sys}}}{G} = \frac{T^*_{\text{sys}}}{\text{DPFU} \times g(\text{el})}$$  \hspace{1cm} (A.53)

It is important to note that the SEFD contains corrections for system noise, atmospheric absorption, antenna gain terms and temperature-to-Jansky conversion.

**A brief overview of a priori amplitude calibration**

For VLBI observations, there are very few suitable calibrators that do not become resolved on some baselines, thus we cannot use the primary calibrator scaling to calibrate VLBI amplitudes. An alternative approach is to calibrate the VLBI amplitudes using the system temperatures and collecting areas of the individual antennas. The visibility amplitudes can be calibrated in units of flux density by multiplying the normalized visibility amplitudes by the geometric mean of the SEFDs of the two antennas concerned (TMS Section 10.1.). On a baseline between two telescopes, for example the SMT and the LMT, which both use the chopper method, the amplitude calibration for the correlated source signal $r_{\text{corr,SMT-LMT}}$ (compensated for digitization and sampling losses) on that baseline is given by:

---

where \( \text{SEFD}_{\text{SMT}} \) and \( \text{SEFD}_{\text{LMT}} \) are determined as shown above and \( S_{\text{SMT-LMT}} \) is then the source signal in Jansky on that baseline.

Since the SEFDs for the telescopes are expected to include the effective system noise temperature, which corresponds to a signal plane above the atmosphere, then the resulting visibility amplitudes will be corrected for atmospheric losses.

### Double-sideband (DSB) receivers

It is worth noting that the equations presented in the previous sections for amplitude calibration are modeled after the SMT, which has a sideband-separating receiver. However, a few stations in the Event Horizon Telescope array have double-sideband (DSB) receivers, which lead to some modifications of the equations for amplitude calibration. The most relevant difference between SSB and DSB receiver is the handling of measured signals. For an SSB receiver, all the measured signal comes from only one sideband, but for a DSB receiver it comes from two sidebands folded together into one single larger band, usually used for spectral-line observing. However, for continuum VLBI with the EHT, only one sideband of the DSB receiver systems is used as the signal sideband and gets correlated, but the rest of the telescope continues to operate as a DSB system. Therefore, the sensitivity of the measurements during EHT observing (through one sideband) is about a factor of two lower than the normal operation of the telescope as a perfect DSB system.

This rescaling of the telescope sensitivity from two sidebands to one is done by correcting \( T^{\ast}_{\text{sys}} \).

For a measured effective system temperature from a perfect DSB system \( T^{\ast}_{\text{sys,DSB}} \), the actual effective system temperature for VLBI observing with only one sideband is:

\[
T^{\ast}_{\text{sys}} = 2T^{\ast}_{\text{sys,DSB}}
\]

(A.55)

For EHT observing we use half the number of sidebands, thus the telescope sensitivity must drop by a factor of two, leading to the effective system temperature increasing by the same factor. However, if the telescope does not have a perfect DSB system but one sideband has more gain than the other, then the equation becomes, more generally:

\[
T^{\ast}_{\text{sys}} = (1 + r_{sb})T^{\ast}_{\text{sys,DSB}},
\]

(A.56)

where the sideband ratio \( r_{sb} \) is the ratio of source signal power in the remaining sideband to the signal power in the sideband of interest (the sideband to be correlated). For a perfect DSB system, the gains of each sideband are equal, giving \( r_{sb} = 1 \), which gives back Eq. A.55. For a perfect SSB system, where all signal is in one sideband, \( r_{sb} = 0 \) and this gives back simply \( T^{\ast}_{\text{sys}} \) needed for the EHT.

Once this correction is applied to \( T^{\ast}_{\text{sys}} \), the rest of the amplitude calibration process remains the same. For planet scans to determine the telescope’s DPFU, the signal is collected by both sidebands in a DSB system, thus the effective antenna temperature is usually measured in DSB mode. This is sufficient to reflect the conversion from Kelvin to Jansky within the aperture.
Appendix A: Amplitude Calibration Memo

efficiency and DPFU estimation. It should be noted that the correction from a DSB system to an SSB system for VLBI should only be done on $T^*_{\text{sys}}$, otherwise the resulting SEFDs would be double-corrected for a DSB system.

A.2.2 Telescopes not using the chopper technique

As explained above, the result for telescopes like the SMT and the LMT, which both use the chopper (or single-load) technique, is very clean and simple. Now what happens when there is a telescope in the array that does not use the chopper technique but instead uses the direct (or noise diode) method?\footnote{As far as the EHT is concerned, there are no stations in the array at this time without the chopper technique. However, this information could be potentially useful for the a priori calibration of GMVA or HSA observations related to EHT, which are a mixture of mm- and cm-observatories.}

In that case, on baselines with telescopes with the chopper method, there will be inconsistencies in the amplitude calibration if the same corrections are applied in post-processing to both stations on that particular baseline. This is precisely because the chopper technique gives $T^*_{\text{sys}}$ and the direct method only gives $T_{\text{sys}}$.

Fortunately, as explained in the previous section, the relationship between the two is well-understood and $T^*_{\text{sys}}$ can easily be determined from the direct method using opacity measurements. If the telescope has a tipping radiometer or water vapor radiometer nearby measuring opacities, this can give a fairly good estimate for $T^*_{\text{sys}} = e^\tau T_{\text{sys}}$.

However, some aspects of radiometers hinder this approach:

- The radiometer does not always point in the same direction as the telescope, thus under a varying or partly cloudy sky the opacities from the radiometer are not entirely accurate to the observations.
- The radiometer can have something blocking and corrupting the measurements (as on Mt Graham due to the LBT)
- The radiometer does not always measure an opacity at the observing frequency but instead is converted (sometimes not so accurately) from a different frequency

Another possible solution is to use the telescope itself as a tipper: using the dish to observe blank sky through a big elevation range in the direction of observing to determine the relationship between elevation and sky temperature and get an estimate of the zenith opacity.

This tipping method solves the radiometer issues of getting an opacity in the direction of observing and at the right observing frequency. However, these tipping scans are required very frequently, every 10 minutes or so, and take up valuable observing time just to get accurate opacities.

An alternative is then to obtain opacities using approximations in post-processing. The system noise temperature, as measured in the direct method, is defined as seen previously. For $\tau_0 \ll 1$, we can approximate:

$$T_{\text{sys}} \approx T_{\text{rx}} + T_{\text{atm}}(1 - e^{-\tau}) \approx T_{\text{rx}} + T_{\text{atm}} \times \tau_0 \times \text{AM}$$  \hspace{1cm} (A.57)
By fitting a least-squares (or as it is done for the GMVA, a linear fit to the lower envelope, Martí-Vidal et al. 2012) of $T_{\text{sys}}$ as a function of airmass, the extrapolation of the fit will give an approximation for the receiver noise temperature $T_{\text{rx}}$. If the telescope frequently does a dual-load (cold cal) calibration to refresh values for the receiver temperature, these values are usually more accurate to use.

With this linear relationship (or measured $T_{\text{rx}}$) and every variable but the sky opacity known, measured or approximated by the telescope, we can get the sky opacity at the zenith and thus correct the system noise temperature for the atmospheric attenuation:

$$
\tau_0 = -\frac{1}{\text{AM}} \ln \left( 1 - \frac{T_{\text{sys}} - T_{\text{rx}}}{T_{\text{atm}}} \right) \quad (A.58)
$$

### A.2.3 $T_{\text{sys}}$ or $T_{\text{sys}}^*$?

A crucial part of the amplitude calibration process is to determine which variables are actually provided by each telescope in the context of the entire EHT array. Are all telescopes providing $T_{\text{sys}}$ like cm-observatories do? Or are some telescopes providing in fact $T_{\text{sys}}^*$ but labeling it as $T_{\text{sys}}$ (as is commonly done by mm-observatories)? Discrepancies in notation and a heavy background knowledge in the context of cm-observatories can cause misunderstandings of the calibration information provided by the telescopes. However, there is a nice way to do a quick check in post-processing.

In order to visually understand the difference between the two variables, simulated measurements of system noise temperatures for the SMT are presented. Using the standard chopper equation, the calibration temperature was approximated to $T_{\text{cal}} = T_{\text{amb}} = 280$K, the receiver noise temperature was set to $T_{\text{rx}} = 60$K, and we have used a constant zenith opacity of $\tau_0 = 0.2$, common for the SMT, for consistency. Figure A.3 shows the effective system noise temperature $T_{\text{sys}}^*$ using the chopper technique equation and the direct method system noise temperature $T_{\text{sys}} \approx e^{-\tau T_{\text{sys}}^*}$ as a function of airmass. It is clear that both temperatures indeed do vary with airmass, but $T_{\text{sys}}^*$ is a lot more sensitive because in addition it corrects for the increasing attenuation of a signal from outside the atmosphere, automatically determined with the chopper technique.

It is a misconception to assume that because $T_{\text{sys}}$ does not contain that term, it does not vary with airmass. $T_{\text{sys}}$ is inherently dependent on airmass because the sky brightness temperature $T_{\text{sky}}$, representing atmospheric noise, increases with airmass as the telescope looks through a larger layer of atmosphere. This effect is also present in $T_{\text{sys}}^*$, which, in addition, corrects for the increasing signal attenuation that is also elevation (airmass) dependent.

In the previous section, we introduced a useful tool to tell the two system noise temperatures apart. For the case of $T_{\text{sys}}$, as shown by eq. A.58, it is possible to untangle a zenith opacity from $T_{\text{sys}}$ and $T_{\text{rx}}$ measurements. In the case of $T_{\text{sys}}^*$, because the opacity relationship is much more complicated, it would not be valid. If we were to apply eq. A.58 using $T_{\text{sys}}^*$ instead of $T_{\text{sys}}$, the opacities at zenith obtained would be highly inaccurate when compared to, for example, radiometer measurements during the same observing window.

---

5This check only works if the telescopes provide an elevation for each “$T_{\text{sys}}$”, or alternatively these can be extracted from the VLBI Monitor database.
This reasoning was thus applied to the SMT to see if what was previously called “$T_{\text{sys}}$” was really $T_{\text{sys}}$. For example, we can take the system noise temperatures and receiver temperature measured during the gain curve measurements for 2017. These were measured in the lapse of a few hours, thus minimizing opacity fluctuations due to changing weather.

![Simulated system temperatures from the chopper and direct method calibration techniques show divergence as a function of airmass.](image)

**Figure A.3:** Simulated system temperatures from the chopper and direct method calibration techniques show divergence as a function of airmass.

![Zenith opacities obtained by assuming the telescope provides $T_{\text{sys}}$ are much larger than what is actually measured by the SMT tipper.](image)

**Figure A.4:** Zenith opacities obtained by assuming the telescope provides $T_{\text{sys}}$ are much larger than what is actually measured by the SMT tipper.

We use the opacity equation:

$$\tau_0 = -\frac{1}{\text{AM}} \ln \left(1 - \frac{T_{\text{sys}} - T_{\text{rx}}}{T_{\text{atm}}} \right)$$

(A.59)

Recall that at this point, what is plugged in as $T_{\text{sys}}$ is what is measured by the chopper technique (in fact it is $T_{\text{sys}}^*$ but this was still unknown).
The zenith opacities obtained from that equation were then compared to the measured zenith opacities by the tipping radiometer on the telescope scan-by-scan. These results are presented in Fig. A.4. It is clear that the zenith opacities obtained from “\(T_{\text{sys}}\)” are completely different, incredibly high and inconsistent with the tipper measurements. This is of course because “\(T_{\text{sys}}\)” is in fact \(T_{\text{sys}}^*\), which diverges and is increasingly larger than \(T_{\text{sys}}\) as a function of airmass. Thus the zenith opacity equation does not work for what is outputted by the chopper technique at the SMT, and this output is definitely not simply \(T_{\text{sys}}\) but something much more sensitive to opacity: \(T_{\text{sys}}^*\). For a telescope that is genuinely providing \(T_{\text{sys}}\), the opacity equation would give results much more in-line with the measured opacities from its tipper/radiometer.
Abstract

This document presents a step-by-step explanation of the Submillimeter Telescope (SMT) calibration procedure for the antenna-based a priori amplitude calibration as part of the Event Horizon Telescope (EHT). During the EHT+ALMA April 2017 observing run, a number of calibration observations and tests were done. The measurement and reduction processes for each of the SMT a priori calibration deliverables are described in this memo. Improvements to the calibration procedure include: a newly estimated and updated beam-width for the telescope at 228.1 GHz enabling a more accurate estimation of the flux density scaling using planet calibrator observations; a step-by-step outline of the system temperature calibration process using exact equations from the telescope software scripts; tests of various parts of the signal chain for potential amplitude losses; and a newly determined gain curve for 2017.

Note: This document can be used in conjunction with similar calibration outlines from other stations for procedural comparisons. It contains the most accurate calibration information for the SMT to date and thus renders results from previous memos obsolete.
Relevant terminology

Relevant variables introduced in this document (brightness temperatures approximated with the Rayleigh-Jeans approximation):

- $C_{\text{hot}}$: Counts measured when looking at the hot load (vane)
- $C_{\text{cold}}$: Counts measured when looking at the cold load (liquid nitrogen)
- $C_{\text{on}}$: Counts measured observing a target
- $C_{\text{sky}}$: Counts measured when looking at blank sky
- $T_{\text{cold}}$: temperature of the cold load
- $T_{\text{rx}}$: receiver noise temperature
- $T_{\text{vane}}$: temperature of the vane (blocker)
- $T_{\text{amb}}$: ambient temperature around the SMT, as measured by the weather station on Mount Graham (physical temperature)
- $T_{\text{em}}$: temperature of the atmospheric emission (also interchangeably called $T_{\text{sky}}$ in the literature, the brightness temperature of the sky)
- $T_{\text{em,i}}$: temperature of the atmospheric emission in the image band (sideband not used for observing)
- $T_{\text{em,s}}$: temperature of the atmospheric emission in the signal band
- $T_{\text{cab}}$: physical temperature of the receiver cabin (this is assumed to be the same as the ambient temperature)
- $T_{\text{cal}}$: derived temperature to give a correct temperature scale for the signal band
- $T_{\text{sys}}$: system noise temperature of the system
- $T^*_{\text{sys}}$: effective system noise temperature (corrected for atmospheric attenuation and rearward losses)

Efficiency and correction terms:

- $r_{\text{sb}}$: sideband ratio - since the SMT is a single-sideband receiver, $r_{\text{sb}} = \frac{g_i}{g_s} \ll 1$ since no signal comes from the image band but some leakage can still be present
- AM: amount of airmass in the line of sight of the receiver (elevation-dependent)
- $\tau_0$: atmospheric opacity at the zenith
- $e^{-\tau}$: atmospheric attenuation factor, which damps the signal based on atmospheric opacity in the line of sight $\tau = \tau_0 \times \text{AM}$
B.1 Introduction

A telescope’s system-equivalent flux density (SEFD) is simply the noise contribution of the system, given by the system noise temperature, and all losses and gains, converted to a flux density scale. The SEFDs can be calculated using system noise temperature $T_{sys}$ measurements and all efficiencies and contributions to source attenuation and noise, and one can determine the sensitivity of the telescope when compared to other telescopes in the array. The higher a telescope’s SEFD, the lower its sensitivity. Ultimately, the flux density of a source is simply the telescope’s SEFD, which contains all system and telescope parameters and efficiencies, multiplied by the ratio of signal to noise power (defined as $r_{S/N}$) of the source detection. The equation for the SEFD can be subdivided into three main components, each with station-based variations for how they are determined and measured. The three components to the SEFD are:

1. $T_{sys}$: the total noise characterization of the system, given by the system noise temperature
2. $\eta_{\tau l}$: the correction terms for attenuation of the source signal by the atmosphere and rearward losses
3. $G$: The antenna gain, including all the loss terms from the telescope and the conversion from a temperature scale (K) to a flux density scale (Jansky), given by the *degrees per flux
density unit' factor (DPFU) in K/Jy and the normalized elevation-dependent gain curve \( g(\text{el}) \):

\[
\text{DPFU} = \frac{\eta_A A_{\text{geom}}}{2k}
\]

, giving antenna gain \( G = \text{DPFU} \times g(\text{el}) \)

This gives the following general equation for a telescope’s SEFD:

\[
\text{SEFD} = \frac{T_{\text{sys}} e^T}{\eta G}
\]

(B.1)

The flux density of a source detected with a given ratio of signal to noise power \( r_{S/N} \) is then:

\[
S_{\text{source}} = \text{SEFD} \times r_{S/N} = \frac{r_{S/N} \times T_{\text{sys}} e^T}{\eta G}
\]

(B.2)

For mm-observatories, which measure the effective system temperature \( T_{\text{sys}}^* = T_{\text{sys}} e^{\tau} \) directly using the chopper technique\(^1\), the SEFD equation can be rewritten in only two components, the effective system noise temperature and the antenna gain [for more information see Kraus, 1966; Burke & Graham-Smith, 2009; Wilson et al., 2013; Thompson et al., 2017]:

\[
\text{SEFD} = \frac{T_{\text{sys}}^*}{G}
\]

(B.3)

For the SMT, the SEFD at zenith is of order 13 000 Jansky.

In order to fully calibrate the SMT amplitude scale, we must have a comprehensive understanding of the variables involved in the calculation of source flux densities and how they are measured and obtained. The following equations and descriptions explain the entire procedure for the SMT. In order to easily understand the variables within LinuxPops, the SMT in-house calibration software, a compilation of each variable in the theoretical calibration below and its LinuxPops counterpart can be found in Appendix ??.

### B.2 Telescope-specific parameters and efficiencies

The most basic parameter relevant for the sensitivity of a parabolic antenna is its geometrical collecting area, which affects the amount of radiation reflected from the primary onto the secondary reflector. The geometric area of the SMT is calculated with a known diameter of the dish of 10 meters:

\[
A_{\text{geom}} = \frac{\pi D^2}{4} \approx 78.54 \, \text{m}^2
\]

(B.4)

The aperture efficiency represents the efficiency of the telescope compared to a telescope with a perfect collecting area (uniform illumination, no blockage or surface errors) and it is determined using observations of known calibrator sources, usually planets. It corrects for the forward losses of the telescope during observing. We can write the aperture efficiency as the combination of various efficiency terms [Baars, 1973; Gordon et al., 1992; Greve & Bremer, 2010]:

\[
\eta_A = \eta_{\text{taper}} \times \eta_{\text{block}} \times \eta_{\text{spillover}} \times \eta_{\text{Ruze}}
\]

(B.5)

Each efficiency term corresponds to an aspect of the telescope feed:\(^2\)

---

\(^1\)See the complementary *A priori Calibration Memo* by Issaoun et al. (2017) for details

\(^2\)See Baars, J., *The paraboloidal reflector antenna in radio astronomy*, Springer, 2007 for more detail on the measurement of the different losses.
B.3 The SMT DPFU

- $\eta_{\text{taper}}$ is the efficiency loss due to non-uniform illumination of the aperture plane by the tapered radiation pattern/feed function (also formally known as the illumination efficiency). It is the most important contributor to the aperture efficiency.

- $\eta_{\text{block}}$ is the aperture blockage efficiency due to blocking of the feed by the sub-reflector (including its support legs)

- $\eta_{\text{spillover}}$ is the feed spillover efficiency past the main reflector - it is the ratio of the power intercepted by the reflective elements beyond the edge of the primary and the sub-reflector to the total power. It is due partly to cold sky and partly to a warm background, and is elevation-dependent.

- $\eta_{\text{Ruze}}$ is the surface error efficiency (also called "Ruze loss" or scattering efficiency) calculated from Ruze’s formula [Ruze, 1952]. It is due to randomly distributed small-scale deviations of the reflector from the perfect paraboloidal shape. Ruze’s formula is presented below, where $\sigma$ is the surface rms (accounting for small-scale deviations from a perfect surface through dish holography) and $\lambda$ is the observing wavelength:

$$\eta_{\text{Ruze}}(\lambda) = e^{-\frac{16\sigma^2\lambda^2}{\lambda^2}}$$

Additionally, telescopes do not have perfect surfaces, and must thus suffer some losses and gains of signal due to distorted illumination of the main reflector as they slowly move to different elevations. This large-scale surface deformation affects the received signal and is not taken into account in the general aperture efficiency calculation. These losses however can be determined by tracking sources through a wide range of elevations, and thus measure an elevation-dependent gain curve for the telescope, where the maximum ($g = 1$) is set where the telescope is expected to be the most efficient. This normalized gain curve is usually written in the form of a second order polynomial (in the standard VLBA format), where ‘el’ is the elevation in degrees:

$$g(\text{el}) = a_2(\text{el})^2 + a_1(\text{el}) + a_0$$  \hspace{1cm} (B.7)

An updated gain curve for 2017 can be found in Section B.8.

B.3 The SMT DPFU

We have previously defined the antenna gain as a combination of the telescope’s DPFU and its normalized elevation-dependent gain curve. To obtain the SMT’s Jy/K factor $f_{\text{Jy/K}}$, independent of elevation, we simply take the inverse of the DPFU:

$$\text{DPFU} = \frac{A_{\text{geom}}\eta_A}{2k} \frac{\text{K}}{\text{Jy}}$$  \hspace{1cm} (B.8)

$$f_{\text{Jy/K}} = \frac{1}{\text{DPFU}} = \frac{2k}{A_{\text{geom}}\eta_A} \frac{\text{Jy}}{\text{K}}$$  \hspace{1cm} (B.9)

The antenna gain will of course also be affected by the gain curve, not included in $f_{\text{Jy/K}}$, depending on the elevation of the observed source. A time-dependent DPFU can also be estimated using flux measurements of planets at different times during the observing window.
B.3.1 The SMT beam-width

In order to confidently provide an aperture efficiency for the telescope, a reliable beam-width must be used. The expected flux densities from calibrators are dependent on the coupling of the source size to the telescope beam. Therefore an accurate estimation of the telescope beam-width is a necessary input for the simulation programs giving expected calibrator fluxes.

After the EHT+ALMA run in April 2017, we made a continuum map of Jupiter to estimate the telescope beam-width at the EHT observing frequency (228.1 GHz). The map is a \(5' \times 5'\) azimuth-elevation (az-el) grid map, with the telescope moving horizontally in azimuth at each elevation increment of \(10''\). The map is shown in Fig. B.1. The SMT has a chopping beam with a separation of \(4'\) between the two beams: during the az-el grid mapping, both the negative and positive beam slew through the map area, hence the two images of Jupiter.

A first attempt at obtaining the beam-width of the telescope was made by fitting a Gaussian to the flux density collected by the telescope at the zero elevation change mark, along the azimuth change. The deconvolution of the beam-width \(\theta_{mb}\) (the width of the main lobe of the telescope beam pattern) from the full width at half-maximum (FWHM) of a Gaussian is estimated as follows (accurate to about 2\% for an extended source). The diameter of Jupiter \(\theta_{Jupiter} = 41.17''\), as computed from the Planet program for 11 April 2017, was used for the deconvolution. The beam-width is thus given by:

\[
\theta_{mb} = \sqrt{\text{FWHM}^2 - \frac{\ln(2)}{2} \theta_{Jupiter}^2}
\]  

This method gave wildly varying beam-widths between the positive and negative beam. Therefore, a different analysis was used for a more robust result: using a least-squares fit of a 2D Gaus-
Figure B.2: Left: Map of Jupiter by the positive beam of the SMT. Center: Best-fit 2D Gaussian model of the positive beam map. Right: Residual map of the model-subtracted image.

Figure B.3: Left: Map of Jupiter by the negative beam of the SMT. Center: Best-fit 2D Gaussian model of the negative beam map. Right: Residual map of the model-subtracted image.
Table B.1: Measured beam-widths from the Jupiter az-el grid map using 2D Gaussian fits for the positive and negative beam.

<table>
<thead>
<tr>
<th>Beam</th>
<th>Beam-width along ∆Az [arcseconds]</th>
<th>Beam-width along ∆El [arcseconds]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>36.1 ± 0.3</td>
<td>34.9 ± 0.3</td>
</tr>
<tr>
<td>Negative</td>
<td>32.6 ± 0.3</td>
<td>35.4 ± 0.3</td>
</tr>
</tbody>
</table>

A 2D Gaussian model was fitted to both the negative and positive beam images of Jupiter, and then the telescope beam size was deconvolved from the resulting FWHM in the azimuth and elevation directions and the known diameter of Jupiter on April 11, 2017, as previously outlined in Eq. B.10. Fig. B.2 and Fig. B.3 show the best-fit 2D Gaussian model for the Jupiter images taken by the positive and negative beams of the SMT respectively. Table B.1 presents the dimensions of the telescope beams estimated from the 2D Gaussian fits.

Of course this method only gives a simple fit, but this should be sufficient for the analysis of the telescope DPFU, in particular because other uncertainties dominate (pointing, focus). The difference between the sizes of the positive and negative beams are not yet fully understood, but it is possibly caused either by technical differences in the chopping of the secondary between the two beams or by non-uniform illumination of the dish from distortions on the surface of the main reflector. The latter is also further shown by the non-symmetrical distortions caused by the secondary support legs, as shown in Fig. B.4.

This is also supported by the fact that the residuals of the model-subtracted images from the two beams are not mirror images of each other but are both biased to the left side: this again points toward asymmetry in dish illumination. Asymmetry between the two beams is also not an uncommon phenomenon for radio-telescopes with the same chopping setup. Other possible effects could be atmospheric distortions, tracking errors or gridding problems during the mapping (the az-el grid map mode is not used frequently at the SMT).

From Table B.1 we can estimate an average beam-width for the SMT at 228.1 GHz: \( \theta_{\text{SMT}} = 34.8 \pm 0.6 \) arcseconds. It is important to note however that an approximation of a circular beam might not be entirely accurate for the SMT, as Table B.1 shows. However, any uncertainty from the beam-width used will not be the dominating factor for the uncertainty in the DPFU, making the average estimate an adequate value to use for the calibration procedure.

B.3.2 The aperture efficiency

With the newly updated telescope beam-width, the aperture efficiency for the SMT can now be estimated from planet observations. We have decided not to use the SMT in-house Planet simulation software for the following calculations (as was usually done in past memos) but to instead only use the brightness temperatures and apparent sizes given for Saturn, Jupiter and Mars and calculate every step of the calibration independently. The newly estimated aperture efficiencies with the method described here render measurements in past memos outdated and inaccurate.
The observed planet fluxes from the SMT calibrator scans are compared to expected planet flux densities in the telescope beam for a perfect telescope at the given frequency and beam-width. The aperture efficiency $\eta_A$ is found using the following equation, where $T_A^*$ is the observed effective antenna temperature, $g(\text{el})$ is the telescope gain curve, $k$ is the Boltzmann constant, $A_{\text{geom}}$ is the geometric area of the telescope and $S_{\text{beam, sim}}$ is the expected flux density of the planet in the telescope beam estimated from planet information such as apparent size and brightness temperature:

$$\eta_A = \frac{2k}{A_{\text{geom}} } \frac{T_A^*}{g(\text{el})S_{\text{beam, sim}}}$$  \hspace{1cm} (B.11)

Or similarly the DPFU is directly given by:

$$\text{DPFU} = \frac{T_A^*}{g(\text{el})S_{\text{beam, sim}}} \hspace{1cm} (B.12)$$

For extended sources, it is important to calibrate the flux density observed in the beam because some emission might not be picked up by the telescope. The aperture efficiency is only concerned by the main beam flux density, and so the following equation is used to calibrate the simulated flux density in the beam for an extended source, where $S_{\text{sim}}$ is the expected total flux density of the source:

$$S_{\text{beam, sim}} = S_{\text{sim}} \times K$$  \hspace{1cm} (B.13)

$K$ is determined by the following equation, where $\theta_{\text{mb}}$ is the half-power beam-width in arcseconds of the primary lobe of the telescope beam pattern (telescope beam diameter) and $\theta_s$ is the diameter in arcseconds of the observed extended source:

$$x = \frac{\theta_s}{\theta_{\text{mb}}} \sqrt{\ln(2)}$$  \hspace{1cm} (B.14)
Table B.2: Typical planet apparent sizes, brightness temperatures and $K$ correction factors for EHT observing at the SMT over the past three years (for Mars, April 2016 only).

<table>
<thead>
<tr>
<th>Planet</th>
<th>$\theta_s$ [arcseconds]</th>
<th>$T_B$ [K]</th>
<th>$K$ factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mars</td>
<td>12 ± 1</td>
<td>208</td>
<td>0.96</td>
</tr>
<tr>
<td>Jupiter</td>
<td>42 ± 3</td>
<td>170</td>
<td>0.63</td>
</tr>
<tr>
<td>Saturn</td>
<td>16 ± 2</td>
<td>150</td>
<td>0.93</td>
</tr>
</tbody>
</table>

\[ K = \frac{1 - e^{-x^2}}{x^2} \] \hspace{1cm} (B.15)

This $K$ factor is the ratio of the beam-weighted source solid angle and the solid angle of the source on the sky. It is in fact the integral of the antenna pattern of the telescope (approximated as a normalized gaussian) $P(\theta, \phi) = e^{-\ln 2(2\theta/\theta_{mb})^2}$ and a disklike source with a uniform brightness distribution $\Psi(\theta, \phi) = 1$ over the size of the extended source. This serves very well for our a priori calibration purposes.

\[ K = \frac{\Omega_{\text{sum}}}{\Omega_s} = \frac{1}{\Omega_s} \int_{\text{source}} P(\theta - \theta', \phi - \phi')\Psi(\theta', \phi')d\Omega' \] \hspace{1cm} (B.16)

\[ K = \frac{1}{\Omega_s} \int_{\text{source}} P(\theta - \theta', \phi - \phi')d\Omega' \] \hspace{1cm} (B.17)

The expected total flux density is estimated by:

\[ S_{\text{sim}} = \frac{2h}{c^2} \frac{\nu^3 \Omega_s}{e^{\frac{h\nu}{kT_B}} - 1}, \] \hspace{1cm} (B.18)

where $\nu$ is the observing frequency in Hz, $h$ is the Planck constant, $c$ is the speed of light (in m/s), $T_B$ is the mean brightness temperature for the planet (assuming a disk of uniform temperature), and $\Omega_s$ is the solid angle of the source on the sky in steradians. Since we are dealing with very small objects, the latter can be approximated using the small angle approximation, where $\theta_s$ is the apparent diameter of the planet observed, in radians:

\[ \Omega_s \approx \frac{\pi \theta_s^2}{4} \] \hspace{1cm} (B.19)

Of course this process heavily depends on assumptions made in the planet calibration, such as accurate predicted planet fluxes, stable weather conditions and a well-calibrated instrument in terms of pointing and focus [Bensch et al., 2001].

For the SMT, the Planet program provided brightness temperatures $T_B$ and apparent angular sizes of each planet at a given date and frequency. The program gave, for each planet and each day, a major and minor axis estimation for the planet disk. We approximated the diameter of the planet $\theta_s$ for the flux density calculation as the mean of the major and minor diameters. Table B.2 shows typical parameters of planet calibrators for EHT observing at the SMT.

Planet scans during three EHT observing windows, in 2015, 2016 and 2017, were taken to estimate an aperture efficiency per scan. The measurements with Saturn were discarded due to the Saturn rings influencing the disk approximation for the planet and the need for a more detailed
B.3 The SMT DPFU

Figure B.5: Aperture efficiency estimates for upper sideband measurements (USB) using Jupiter and Mars (2015-2017). The blue dashed line represents the mean aperture efficiency from the measurements. The error bars are estimated from the standard deviation from the mean of the measurements approximated as a Gaussian distribution.

brightness distribution model for accurate estimates of an aperture efficiency. Furthermore, only antenna temperature measurements done in upper sideband (USB) were kept for the analysis, as this is the sideband used for EHT observing. Separate aperture efficiency measurements were also determined per polarization (RCP and LCP). Fig. B.5 shows the measurements used for the analysis and the resulting mean aperture efficiency for the SMT.

An attempt was also made to determine a time-dependent variation of the aperture efficiency as a function of UT time, as shown in Fig. B.6. A time-dependence of the aperture efficiency can be caused by the dish being affected by the Sun and heating up during daytime hours, thus reducing telescope sensitivity and efficiency. However, a very peculiar physical trend was observed at the SMT as a function of UT time (local time is in MST = UT - 7). It appears that the telescope is at its most efficient during mid-day and reaches a minimum in the night. The current measurements, which appear to show an increase in efficiency during daytime hours, are not trusted.

The lack of a trusted physical trend for the time-dependence of the aperture efficiency thus led to the conclusion that a single average aperture efficiency must be adopted for the SMT until this behavior can be better understood or a better sampling of planet measurements (not just during EHT weeks in April) can be obtained. The telescope is not particularly known to be under-performing or having difficulties observing in daytime conditions.

The aperture efficiency of the SMT is determined by taking the mean of all aperture efficiency measurements from all three EHT observing windows, and the error on the measurement is the standard deviation determined from the scatter of the measurements around the mean value. It is likely that the error on the aperture efficiency is overestimated, as the scatter of the measurements can be caused by various other effects such as the unsolved time-dependent trend, uncertainties and fluctuations in the measurement of the effective system temperature $T_{\text{sys}}^*$ (used to calibrate
Figure B.6: Aperture efficiency estimates using Jupiter (squares) and Mars (circles). 2015 points in blue, 2016 in green and 2017 in red.

Figure B.7: Aperture efficiency measurements separated by polarization: blue is LCP, red is RCP. Once again, the squares are Jupiter and circles are Mars measurements.
counts on source to $T_A^*$ scale), changes in pointing and focus accuracy or a variable atmosphere due to sparse cloud cover and high winds. In particular, the uncertainty in the $T_{sys}^*$ measurement is also present in the SEFD calculation for the telescope, thus it is likely that its error contribution is double-counted: once for the actual $T_{sys}^*$ values needed to calculate the SEFDs; and once again for their effect on the $\eta_A$ measurements.

The mean aperture efficiency and DPFU (or Kelvin-to-Jansky factor $f_{Jy/K}$) for the SMT are thus\textsuperscript{3}:

$$\eta_A = 0.59 \pm 0.03$$ (B.20)

$$\text{DPFU} = \frac{\eta_A A_{\text{geom}}}{2k} = 0.0168 \pm 0.0008 \text{ [K/Jy]}$$ (B.21)

$$f_{Jy/K} = \frac{2k}{\eta_A A_{\text{geom}}} = 59 \pm 3 \text{ [Jy/K]}$$ (B.22)

Furthermore, the difference between RCP and LCP aperture efficiency (or DPFU) was not found to be significant. This can be seen in Fig. B.7 for individual measurements and in Figs. B.8 and B.9 for the mean distributions of aperture efficiency measurements for RCP and LCP respectively.

\section*{B.4 System noise temperature}

The first step to calibrating the flux density measurements of various sources is to determine the noise temperature of the system when the telescope is pointed off-source (blank sky). At the SMT, this is done with the chopper method, by getting a system-to-background ratio of the signal, comparing the power obtained when the telescope is pointed to the sky directly (off-source counts), and when the receiver is blocked by an absorber (chopper counts), and then calibrated to a temperature scale. The following is an account of the exact equations used in the telescope software to determine the effective system noise temperature and other calibration information.\textsuperscript{4}

A more general approximate procedure can be found in the complementary \textit{A priori Calibration Memo} by Issaoun et al. (2017), see Appendix A.

During a cold load calibration with a liquid nitrogen bath (hereby referred to as “cold cal”), the Y-factor and the receiver noise temperature are computed. This is done once every few hours by the operator to refresh receiver noise temperature values. They do not vary much between cold cals:

$$Y = \frac{C_{\text{hot}}}{C_{\text{cold}}}$$ (B.23)

where the numerator is $C_{\text{hot}}$, the counts obtained from the vane (chopper), and the denominator is $C_{\text{cold}}$, the counts obtained from the cold load (liquid nitrogen bath). The Y-factor also enables an easy diagnostic of the sensitivity of the receiver. A high Y-factor means little receiver noise, and thus sensitive observations (of course what constitutes “high” depends on the type of receiver

\textsuperscript{3}More digits for the measurements can be obtained upon request.

\textsuperscript{4}Note: The SMT calibration script written by Thomas Folkers uses the ATM program, a sky emissivity correlation model of sky temperature part of the ASTRO package of the GILDAS software by IRAM. For more detail on this procedure, see \textit{Calibration of spectral line data at the IRAM 30m radio telescope} by C. Kramer.
Figure B.8: Aperture efficiency distribution for RCP approximated by a Gaussian fit of a mean $\eta_{A,RCP} = 0.5911$ and a standard deviation of $\sigma_{RCP} = 0.0285$.

Figure B.9: Aperture efficiency distribution for LCP approximated by a Gaussian fit of a mean $\eta_{A,LCP} = 0.5904$ and a standard deviation of $\sigma_{LCP} = 0.0281$. 
and the observing frequency). For the SMT, the Y-factor is typically \( \sim 2.5 \) for EHT observing. Then the cold cal routine calculates the receiver noise temperature as follows:

\[
T_{\text{rx}} = \frac{T_{\text{vane}} - YT_{\text{cold}}}{Y - 1} \tag{B.24}
\]

Here temperatures are used, where \( T_{\text{vane}} \) is the temperature of the vane blocker (chopper; at room temperature \( \sim 290 \) K) and \( T_{\text{cold}} \) is the cold load temperature (for the SMT it is the temperature of liquid nitrogen \( \sim 77 \) K).

Then the routine computes an estimate for \( T_{\text{em}} = T_{\text{sky}} \), the atmospheric emission temperature (or equivalently the sky brightness temperature), which is later corrected in an iterative process with the ATM program to determine \( T_{\text{sys}}^* \):

\[
T_{\text{em}} = T_{\text{vane}} - \frac{C_{\text{hot}} - C_{\text{sky}}}{C_{\text{hot}} - C_{\text{cold}}} (T_{\text{vane}} - T_{\text{cold}}) \tag{B.25}
\]

Here \( C_{\text{sky}} \) corresponds to the counts obtained by looking at blank sky. When the calibration routine is not a cold cal (but is instead a \( T_{\text{sys}}^* \) routine before a scan/pointing/focus), the program uses the following to estimate \( T_{\text{em}} \):

\[
T_{\text{em}} = (T_{\text{rx}} + T_{\text{vane}}) \frac{C_{\text{sky}}}{C_{\text{hot}}} - T_{\text{rx}} \tag{B.26}
\]

The program then iteratively corrects \( T_{\text{em}} \) using the ATM program to determine the variables in the separate sidebands (signal and image). The atmospheric model (ATM) is used to fit the emission of both sidebands to the sky temperature \( T_{\text{em}} \) by varying the amount of water vapor [Pardo et al., 2001]. The model uses a standard atmosphere and radiative transfer to compute the total absorption and thermal emission by water vapor and oxygen through the atmosphere. After running the ATM transmission routines for the two sidebands, the atmospheric emission temperature for the image band is computed and simplified as follows:

\[
T_{\text{em},i} = T_{\text{em},s} + r_{sb} T_{\text{em},i} \frac{T_{\text{vane}}}{1 + r_{sb}} + T_{\text{cab}} = T_{\text{em},s} \tag{B.27}
\]

With the assumption on the sideband ratio \( r_{sb} = 0 \) for single-sideband (SSB) receiver) and the fact that the cabin temperature is computed as \( T_{\text{cab}} = 0.8T_{\text{vane}} + 0.2T_{\text{amb}} = 0.8T_{\text{amb}} + 0.2T_{\text{amb}} = T_{\text{amb}} \), the atmospheric emission temperatures in the image band and signal band are equal\(^5\). This of course depends on the atmospheric absorption spectrum but remains a fair assumption for the Event Horizon Telescope observing frequencies.

Then the program proceeds to compute the calibration temperature \( T_{\text{cal}} \), which gives the temperature scale for the signal band:

\[
T_{\text{cal}} = (1.0 + r_{sb})(T_{\text{vane}} - T_{\text{em},i})e^{\tau_{0,s} \times AM} = (T_{\text{vane}} - T_{\text{em},i})e^{\tau_{0,s} \times AM} \tag{B.28}
\]

The opacity at the zenith in the signal band \( \tau_{0,s} \) comes from the ATM program, as does the atmospheric emission temperature in the image band \( T_{\text{em},i} \). The program uses elaborate

\(^5\)The assumption that the vane temperature is the same as the ambient/outdoor temperature is no longer trusted. The vane was moved from the receiver cabin to a room in the telescope building, and so the vane temperature is likely closer to room temperature. This is an outstanding issue of the \( T_{\text{sys}}^* \) calibration.
equations to compute the airmass, designed for low-elevation observing. These equations will not be explained here, but they do not deviate from the standard AM = 1/sin(el) equation until below 15° (at 15° it deviates by less than 1%).

The calibration temperature represents the difference between the temperature of the sky and the temperature of the vane, corrected for atmospheric losses, determined by the ATM program. In the standard chopper calibration technique, \( T_{cal} \sim T_{hot} \). It is important to note that the exponential term here is highly misleading. The presence of this term in \( T_{cal} \) serves to stabilize fluctuations in \( T_{cal} \) coming from changes in the atmospheric emission temperature with opacity and elevation. It is NOT the cause of the atmospheric correction factor from the chopper technique resulting in the measurement of \( T^{*}_{sys} \). Actual values for \( T_{cal} \) are very stable, fluctuate only by 1-2%, and do not change as a function of airmass.

Then the program calculates the effective system temperature (including the atmospheric attenuation correction) \( T^{*}_{sys} \) in the following way:

\[
T^{*}_{sys} = \frac{C_{sky}}{C_{hot} - C_{sky}} T_{cal}
\]  

(B.29)

The chopper technique determines a system temperature including all noise contributions from the receiver until the top of the atmosphere [Berdahl & Fromberg, 1982; Berger et al., 1984; Berdahl & Martin, 1984]. Thus, the temperature it outputs is the effective system noise temperature \( T^{*}_{sys} \), which, when applied to source measurements, would already correct for the source signal attenuation by the atmosphere. This is measured before every scan on a target. In VLBI mode, the quarter wave-plate is added to the signal chain, but since \( T^{*}_{sys} \) is measured before every scan in VLBI mode, the chopper calibration is done with the quarter wave-plate in place. Thus any losses induced by the quarter wave-plate would be corrected via the chopper calibration and will be offset within the \( T^{*}_{sys} \) measurement.

### B.5 Obtaining the flux density of a source

We define the antenna temperature for each source measurement, using the signal-to-noise ratio of each measurement, as:

\[
T^{*}_{A} = \frac{C_{on} - C_{sky}}{C_{sky}} T^{*}_{sys}
\]  

(B.30)

Notice the (*) symbol: this is because this antenna temperature is corrected for atmospheric absorption, implicitly included in \( T^{*}_{sys} \).

Now correcting for the aperture efficiency and the gain curve as a function of elevation:

\[
T_{R} = \frac{T^{*}_{A}}{\eta_{A} g(\text{cl})}
\]  

(B.31)

\( T_{R} \) is the final antenna temperature: it contains all corrections for efficiency and telescope parameter terms, and is thought to be the equivalent of the temperature of a resistor held directly in front of the receiver. Thus, it should account for every telescope-specific and source-specific variables, apart from the conversion to a flux density.
B.6 Sky opacity at the SMT

The final step is to convert the temperature scale into a flux density scale in Jansky, where
\[ k = 1.38 \times 10^{-23} \text{J/K} = 1.38 \times 10^3 \text{Jy/K} \]
is the Boltzmann constant:
\[
S = \frac{2k}{A_{\text{geom}}} T_R = \frac{8k}{\pi D^2} \frac{T_\lambda^*}{\eta A \eta_g(A)}
\]  
(B.32)

B.6 Sky opacity at the SMT

The sky opacity for a particular air mass (AM) changes depending on the elevation of the dish as shown below, where \( \tau_0 \) is the sky opacity at the zenith:
\[
\tau = \tau_0 \times \text{AM}
\]
(B.33)

The opacity at the SMT is measured by a tipping radiometer, which is placed about 100° away from the dish, on the telescope building itself. The SMT building is on a rotating platform, therefore the opacity is not measured through a constant water vapor column but rotates to measure 100° from the target source Azimuth.

In particular, this becomes problematic when the telescope observes targets at 200-210° Azimuth, as the tipper is pointed right at the Large Binocular Telescope (LBT). The warm air and reflection from the LBT building causes abnormal spikes in the tipper opacity readings, and these readings become unreliable and not representative of observing conditions.

Furthermore, in order to correct \( T_{\text{cal}} \) for atmospheric attenuation, the ATM program uses an iterative linear fit process to determine an opacity, using weather conditions obtained from the Mount Graham weather station, and can deem a tipper reading as “unreliable” if the measured opacity does not match that outputted by the ATM program. That value is then not used for the calculation of the effective system noise temperature, and it instead uses the last good tipper value to compute this. The ATM program then computes the opacity in the signal band in the line of sight at the exact observing frequency using an iterative process, with the tipper opacity and other parameters from the Mount Graham weather station as inputs.

B.7 Signal loss in the VLBI backend

When calibrating VLBI data using individual telescope parameters and calibration information, there may be discrepancies in the instrumental setups for the observing run and the calibration procedures. Since calibration information is usually obtained via the telescope’s in-house system, there may be some additional unaccounted effects in the recorded VLBI data as it is obtained through the separate VLBI backend (R2DBEs, Mark6 recorders and modules).

In April 2017, we attempted to constrain possible losses through the VLBI backend using a Y-factor test. This was done by measuring and comparing the Y-factor (ratio between a hot/ambient load and a cold load, see Eq. B.23) obtained at the receiver total-power box (part of the SMT setup) and inside the R2DBE itself (R2DBE output6).

A spectral analyzer was used for the total-power box and R2DBE input to compare the results directly with the Y-factor measurements outputted by the R2DBE. It was found that the Y-factors

---

6This was done using the LMT scripts `collect_mark6_data.py` and `genYSFactor.py` written by L. Blackburn and K. Bouman.
at the receiver total-power box and at the VLBI backend, shown for RCP and LCP for both the high and low bands in Fig. B.10, are compatible. It is therefore safe to assume that there is no obvious signal loss through to the R2DBE backend that is unaccounted for in the standard amplitude calibration described in this memo for EHT data taken at the SMT from 2015 onward.

Before 2015, the SMT used the R1DBE backend, where no Y-factor test was done, so it remains difficult to account for possible signal losses in older datasets.

B.8 Updated Gain Curve for 2017

Telescopes are not perfectly rigid paraboloids, and must thus suffer some losses of signal due to distorted illumination of the main reflector as they slowly move to different elevations. This large-scale surface deformation affects the received signal and is not taken into account in the measurements leading to the efficiency and DPFU characterization. These losses can be deter...
mined by tracking sources over a wide elevation range, preferably as they increase and decrease in elevation from the tree-line to transit and vice-versa. An elevation-dependent gain curve for the telescope is measured, where the maximum \((g = 1)\) is set where the telescope is expected to be most efficient. The gain curve is estimated by fitting a polynomial (usually second-order for standard radio-dishes). If more than one source is used, this is done once the flux density measurements are normalized around a plateau (to a relative gain scale). The source measurements used to obtain a gain curve must of course be calibrated for all other effects, including telescope efficiency (through the DPFU) and the atmospheric attenuation of the signal (through \(T^*_{\text{sys}}\)).

Analogously to the previous runs in 2015 and 2016, continuum measurements were made at 1.3 mm of various non-variable calibrator sources. The elevation dependence was investigated using two non-variable sources tracked across the sky for a short period of time (10 hours over one night) and a wide range of elevations: K3-50, a planetary nebula; and W75N, a star formation region [Sandell, 1994]. This tracking was the best method to isolate the elevation-dependent component in the SMT output. The sources reached higher elevations but disappeared at very low elevations behind the solar avoidance zone. The relative gain for the source measurements were taken using the ratio of the measurement against an average value of the upper envelope of data points in a mid-to-high elevation range where the measurements were constant— for a radio telescope, this typically occurs between 40 and 60°. The effect of elevation on the relative gain can be clearly seen in Fig. B.11.

The elevation-dependent geometric (opacity-free) gain curve can be constrained with little-to-no elevation effect from the atmosphere, an effect corrected by the \(T^*_{\text{sys}}\) measurements from the chopper calibration. It is best approximated as the second-order polynomial shown in Fig. B.11. Relative error bars are obtained from the standard deviation of the non-corrected gain distribution around unity [Bevington & Robinson, 2003].

Observing was done in a similar manner to the 2016 gain curve, constraining the measurements to elevations above 20°. The gain curve for the SMT for 2017 is once again a typical radio-telescope elevation gain curve, with a plateau around 40-60°.

The best fit parameters of the elevation-dependent gain curve (of the form \(a_2(\text{el})^2 + a_1(\text{el}) + a_0\)) are shown in Table B.3 below⁷. These are consistent within the error margins with the coefficients of the gain curve for 2016 determined in Memo 2. This result is reassuring, as no work has been done on the telescope in the past year and so the gain curve was not expected to change.

Following the constraint of the elevation gain curve \(g(\text{el})\), LinuxPops can contain the gain curve terms and incorporate them in the SMT calibration procedure in the following way, where

---

⁷More digits and the covariance matrix for the fit parameters can be obtained upon request.
Figure B.11: The best-fit second order polynomial to the relative gains of the two calibrator sources, K3-50 and W75N is used as the elevation-dependent gain curve.

the elevation is in degrees:

\[ g = -0.8 \times 10^{-4}(el)^2 + 0.9 \times 10^{-2}(el) + 0.73 \]  \hspace{1cm} (B.34)

The final flux density is then given by:

\[ S = \frac{1}{\eta_A g(el)} \frac{2k}{A_{\text{geom}}} T^*_{A} \]  \hspace{1cm} (B.35)

B.9 SMT 1.3 mm Measurements

During the EHT+ALMA observing run in April 2017, flux density measurements were obtained for a number of EHT science targets and calibrator sources, by averaging multiple 2.7 min continuum ON-OFF scans in single-dish mode over the EHT observing window (April 4-12, 2017).

The cadence of single-dish observations was not constant or identical between all the sources: some sources were observed more frequently due to their availability during SMT off-time from the VLBI schedules. Table B.4 below presents the mean flux densities for these observed sources during the EHT + ALMA observing run in April 2017. Not all EHT sources are present, some were not visible for long enough during the run to observe them separately with the SMT alone.

B.10 Conclusion

The April 2017 run gave the opportunity to re-evaluate every step of the calibration procedure for amplitude scaling for the SMT. A more detailed understanding of the calibration procedure
Table B.4: Mean flux densities of most EHT observing targets (science targets and calibrators) at 228.1 GHz (1.3 mm) during the EHT+ALMA observing run in April 2017.

<table>
<thead>
<tr>
<th>Source name</th>
<th>Mean flux density [Jy]</th>
<th>± Statistical uncertainty [Jy]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C84</td>
<td>11.4</td>
<td>0.3</td>
</tr>
<tr>
<td>3C273</td>
<td>7.2</td>
<td>0.2</td>
</tr>
<tr>
<td>3C279</td>
<td>8.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3C345</td>
<td>2.3</td>
<td>0.1</td>
</tr>
<tr>
<td>0510+180</td>
<td>2.1</td>
<td>0.3</td>
</tr>
<tr>
<td>BLLAC</td>
<td>1.7</td>
<td>0.1</td>
</tr>
<tr>
<td>CENA</td>
<td>8.5</td>
<td>0.3</td>
</tr>
<tr>
<td>M87</td>
<td>2.1</td>
<td>0.3</td>
</tr>
<tr>
<td>OJ287</td>
<td>3.9</td>
<td>0.3</td>
</tr>
<tr>
<td>SGR A</td>
<td>9.0</td>
<td>0.6</td>
</tr>
</tbody>
</table>

for the telescope was obtained from previous memos, and enabled a more critical evaluation of the strategy currently used. Some limitations were found concerning the telescope DPFU and aperture efficiency: one of the key results of this memo is an updated estimate for the beam-width of 34.8 ± 0.6 arcseconds at 228 GHz, for a more accurate coupling of the planet calibrators to the telescope beam than in previous years. Furthermore, a more realistic estimation of the aperture efficiency was obtained, no differences in DPFU were found between the two polarizations, and measurements were quite consistent over the years, with no significant time-dependence detected.

A test of the signal loss in the VLBI backend was also done in 2017, to investigate the hypothesis that a possible discrepancy between the single-dish calibration information and the scaling needed for EHT data was due to attenuation of signal in the VLBI equipment. It was found, using the Y-factor test at the SMT total-power box, at the input to the R2DBE and within the R2DBE itself, that no signal loss was occurring and the Y-factor results were consistent with each other. It is therefore safe to assume that the single-dish calibration procedure should be suitable to calibrate SMT visibility amplitudes in EHT data in its entirety.

This memo also described the full procedure for the system temperature estimation via the chopper technique, using exact equations from the telescope software written by Tom Folkers. We confirmed that the output of the chopper technique is the effective system temperature $T_{\text{sys}}^*$, which already corrects for atmospheric attenuation of the signal and rearward losses. The $T_{\text{sys}}^*$ measurements are done before each VLBI scan, but the SMT additionally records total-power data that can be used to track intra-scan system temperature trends. This total-power data is available on the EHT wiki as part of the SMT logs for April 2017 observing.

The gain curve was also determined once more in 2017. The resulting fit parameters are consistent, within error estimates, with the curve measured in 2016 (Memo 2). This is an expected result as no work has been done on the telescope since, and it confirms the robustness of the method used to determine the gain curve.

The complete calibration procedure, when taking into account elevation-dependent gain from the gain curve and the newly determined DPFU, yields all the necessary calibration information for
adequate antenna-based scaling of the visibility amplitudes from the SMT for EHT observations. For example, the calibration information is currently consistent with the necessary scaling done by Dr. Rusen Lu and Dr. Thomas Krichbaum for the amplitude calibration of the SMT baselines in EHT 2013 data.

There remains, however, a number of limitations in the calibration procedure that should be investigated further in the future. The beam-width estimation, although adequate considering the uncertainty in pointing and focus dominates the DPFU analysis, is still not quite accurate. We have assumed that the SMT has a circular beam, but from the Gaussian fits it appears to be elliptical instead. A more accurate fit of the Jupiter maps, such as fitting both beams simultaneously or fitting a convolution of the planet disk and a Gaussian instead of a simple Gaussian would yield better results. Another limitation in the calibration is the chopper procedure: the assumption that the vane temperature is the same as the ambient temperature is likely incorrect, as the vane was moved from the receiver cabin (which is at around ambient temperature) to a room in the telescope building (which is closer to room temperature). Some additional assumptions, such as whether the sideband ratio is indeed negligible, must also be looked at more carefully. The chopper technique itself also has its own limitations outside of telescope-specific assumptions. Nevertheless, the current method should yield suitable results for the calibration of EHT observations.

Acknowledgements

This work is supported by the ERC Synergy Grant “BlackHoleCam: Imaging the Event Horizon of Black Holes” (Grant 610058). D.P.M., J.K. and T.W.F. receive support via NSF MSIP award AST-1440254.
Research Data Management

The data presented in this work follow the guidelines of the Data Management Policy of the Institute for Mathematics, Astrophysics, and Particle Physics Research at Radboud University. The data from Chapters 2, 3 and 4 are stored indefinitely on the Global Millimeter VLBI Array and National Radio Astronomy Observatory data archives (https://archive.nrao.edu) under project codes MB007, BG221A, and MJ001. The ALMA interferometric data are available at the ALMA data archive under project codes ADS/JAO.ALMA2016.1.00413.V and ADS/-JAO.ALMA2017.1.00795.V. The raw data used in Chapters 5 and 6 will be stored in the ALMA data archive under project code ADS/JAO.ALMA2016.1.01154.V. Derived data used in the EHT publications are available at the CyVerse Data Commons cloud storage and accessible via the EHT website (https://eventhorizontelescope.org/for-astronomers/data).
Berdahl, P., & Martin, M. 1984, Solar Energy, 32, 663
Berger, X., Buriot, D., & Garnier, F. 1984, Solar Energy, 32, 725
Bevington, P. R., & Robinson, D. K. 2003, Data reduction and error analysis for the physical sciences (McGraw Hill)
Bouman, K. L. 2017, PhD thesis, Massachusetts Institute of Technology
Chirenti, C. B. M. H., & Rezzolla, L. 2007, Classical and Quantum Gravity, 24, 4191

258
Bibliography


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


Einstein, A. 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 844


—. 2019c, ApJL, 875, L3

—. 2019d, ApJL, 875, L4

—. 2019e, ApJL, 875, L5


Greisen, E. W. 2011, AIPS Memo Series, 114r, Socorro, New Mexico, USA
Hu, M.-K. 1962, 8, 179
Iguchi, S. 2005, PASJ, 57, 643
Bibliography

Kraus, J. D. 1966, Radio astronomy
Lüst, R. 1952, Zeitschrift Naturforschung Teil A, 7, 87
—. 2002, ALMA Memo Series, 434
Michell, J. 1784, Philosophical Transactions of the Royal Society of London Series I, 74, 35
Novikov, I. D., & Thorne, K. S. 1973, in Black Holes (Les Astres Occlus), 343–450
Oppenheimer, J. R., & Volkoff, G. M. 1939, Physical Review, 55, 374

262
Ruze, J. 1952, Il Nuovo Cimento, 9, 364
Rybicki, G. B., & Lightman, A. P. 1979, Radiative processes in astrophysics
Schmidt, M. 1963, Nature, 197, 1040
Shepherd, M. 2011, Difmap: Synthesis Imaging of Visibility Data, Astrophysics Source Code Library, , , ascl:1103.001
Smith, F. G. 1951, Nature, 168, 555
Steel, S., Wielgus, M., Blackburn, L., Issaoun, S., & Johnson, M. 2019, EHT Memo Series, 2019-CE-03
Tatarskii, V. I. 1971, The effects of the turbulent atmosphere on wave propagation
Tolman, R. C. 1939, Physical Review, 55, 364
van Cittert, P. H. 1934, Physica, 1, 201
Webster, B. L., & Murdin, P. 1972, Nature, 235, 37
Weizsäcker, C. F. 1948, Zeitschrift Naturforschung Teil A, 3, 524

264
Wielgus, M., Blackburn, L., Issaoun, S., Janssen, M., & Koay, J. 2019, EHT Memo Series, 2019-CE-02
Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2013, Tools of Radio Astronomy, doi:10.1007/978-3-642-39950-3
Zernike, F. 1938, Physica, 5, 785
Supermassive black holes (SMBHs) generate the highest energy processes in the known Universe and eject jets of plasma at near-relativistic speeds affecting galaxy environments on large scales. These objects seem to have symbiotic relationships with their host galaxies, but remain shrouded in mystery. Understanding fundamental properties of black holes and their accretion and emission mechanisms requires a view of the objects across the entire electromagnetic spectrum. Sagittarius A* (Sgr A*) and M87*, the SMBHs at the center of our own Galaxy and the M87 galaxy respectively, are the most promising targets for studies of dynamical processes in extreme gravity across the entire electromagnetic spectrum. My work is integral to the core science of the Event Horizon Telescope (EHT). The EHT project operates a global very long baseline interferometric (VLBI) telescope network at a wavelength of 1mm studying black hole accretion and emission processes, and ultimately testing Einstein’s theory of General Relativity (GR) via direct imaging of black hole shadows.

In Chapter 2, we present the analysis of 3mm observations of Sgr A* with the Global Millimeter VLBI Array (GMVA) joined by the highly sensitive Atacama Large Millimeter Array (ALMA) in 2017, complementary to the EHT. Observations at 3mm are ideal to elucidate the long-standing debate on the nature of the primary radio emission process in Sgr A*, since the image morphology is not dominated by the black hole shadow (as is the case at 1mm) or dominated by effects of scattering in the interstellar plasma (as is the case at longer wavelengths). With the improved resolution and sensitivity gained with ALMA, we reconstructed an intrinsic image of Sgr A* for the first time at 3mm in the first publication at any wavelength to use ALMA as an element of a VLBI array. This work used newly developed imaging techniques such as scattering mitigation and second-moment regularization (Chapter 3) to recover the intrinsic structure of Sgr A*. We demonstrated that the intrinsic image and underlying data tightly constrain a selection of accretion flow models onto Sgr A* to disk- or nearly face-on jet-dominated emission models.

In Chapter 3, we present the second moment regularization technique developed to aid the imaging of compact sources with VLBI. For our GMVA+ALMA observations, it aided the imaging of Sgr A* data, where short-baseline information was dominated by low signal-to-noise measurements driving the images to non-physical intensity distributions. The technique is also a useful tool for imaging EHT data: the imaging fidelity of the EHT is currently determined by its sparse baseline coverage, and is highly sensitive to atmospheric conditions and loss of sites between experiments. Our algorithmic contingency enforces a specific second moment on the reconstructed...
image in the form of a size constraint, and enables the recovery of information lost in gaps of the baseline coverage at short baselines. This second moment regularization technique is a particularly powerful tool for dynamical (movie) reconstructions, where individual snapshots have very sparse coverage, and is currently employed for ongoing imaging work on Sgr A* with the EHT.

In Chapter 4, we present studies of Sgr A*'s intrinsic morphology and interstellar scattering with 2017 and 2018 GMVA+ALMA observations at 3 mm wavelength. Our new 2018 observations confirm non-Gaussian structure in the scattered image seen in 2017, providing a tight constraint on the source size and an upper limit on the dissipation scale of interstellar turbulence. Simultaneously fitting for the scattering parameters, we find an at-most modestly asymmetrical intrinsic source morphology for Sgr A*. In addition to the overall image morphology, comparisons of 2017 and 2018 observations against predictions of scattering models conclusively rule out a model with strong small-scale image distortions, meaning that scattering will only weakly affect 1 mm images with the EHT.

In April 2019 we published the first 1 mm imaging results for the SMBH in the center of the M87 galaxy, revealing a black hole shadow produced by extreme gravitational lensing. In Chapter 5, we present the data processing and calibration tools developed especially to tackle the unique properties of EHT observations, the backbone of the 12 orders of magnitude in data reduction between raw recordings and the now-famous M87* black hole image. In addition to the difficulty of campaign coordination, acquisition and sheer volume of data, the heterogeneity of the EHT array and its susceptibility to weather and atmospheric turbulence at 1 mm make the data calibration particularly arduous. The correlation stage, where the raw telescope data are combined and common signals are detected, is carried out at two central computing facilities. Three independent pipelines were developed to correct for instrumental and atmospheric delays in the arrival time of the signals at the telescopes, building on legacy EHT, low-frequency VLBI, and newly available processing software. Parallel calibration pipelines enabled an extensive suite of cross-validation tests to best quantify data quality and systematics. To disentangle intrinsic source signal from telescope behavior, detailed studies of telescope operations, sensitivities, and observing conditions were carried out. The final calibrated M87* data, ready for scientific analysis and imaging, exhibit clear indications of an asymmetric ring-like structure, with slight structural variations over the course of the observing campaign.

In Chapter 6, we present a subset of the analysis and validation procedures for the total intensity and polarimetric imaging of the M87* black hole, focusing on data and instrument properties as part of the imaging process. Our 2019 total intensity images show a prominent ring with a diameter of ∼40 micro-arcseconds with an enhanced brightness in the southern portion, persistent across four observing nights. To assess the reliability of these results, we implemented a two-stage imaging procedure. In the first stage, four teams, produced blind images of M87* using both the established CLEAN imaging method and newer regularized maximum likelihood (RML) methods. This stage avoided shared human bias by limiting inter-team communication and allowed us to assess common features among independent reconstructions. In the second stage, we reconstructed synthetic data with known ground truth from a large survey of imaging parameters. We selected parameters objectively based on synthetic data fidelity to use when reconstructing images of M87*. The reconstructed station behaviors obtained from the final images were compared to behaviors reconstructed from calibration quasar 3C279, and showed
great consistency as expected. Our 2021 polarimetric images of M87* showed a spiral pattern of the electric vector position angles and a peak linear polarization of 15% concentrated in the south-west portion of the ring. Polarimetric imaging of the M87* black hole followed a similar two-step validation process: the preliminary results were obtained by separating into three software-specific teams, focusing on agreement of derived instrumental leakages for the EHT stations; then a synthetic data survey was carried out to select optimal parameters. The reconstructed station leakages obtained from the final images were compared to those derived from imaging of calibrators as part of our extensive validation tests.

This is only the beginning for EHT science. We are currently preparing the release of horizon-scale results for Sgr A*, which, in conjunction with our significant multi-wavelength partnerships including the 3 mm effort, will provide some insights into very exciting science. The EHT will continue to observe and expand in the coming years in the number of telescopes and to 0.8 mm observing, with a next-generation EHT project in the planning and design stage for the next two decades. With these technical expansions, we hope to make movies of our two main targets and improve imaging fidelity of the black hole shadow for GR tests and of extended structure for jet and accretion studies.
Superzware zwarte gaten (SZG) genereren de meest energetische processen in het bekende heelal en lanceren plasmastralen uit met bijna relativistische snelheden die de omgevingen van sterrenstelsels op grote schaal beïnvloeden. Deze objecten lijken symbiotische relaties te hebben met hun sterrenstelsels, maar blijven in mysterie gehuld. Het begrijpen van fundamentele eigenschappen van zwarte gaten en hun accretie- en emissiemechanismen vereist een zicht op de objecten over het hele elektromagnetische spectrum. Sagittarius A* (Sgr A*) en M87*, de SZG in het centrum van respectievelijk ons eigen Melkwegstelsel en het M87-sterrenstelsel, zijn de meest veelbelovende doelen voor onderzoek naar dynamische processen onder extreme zwaartekracht over het hele elektromagnetische spectrum. Mijn werk maakt integraal deel uit van de kernwetenschap van de Event Horizon Telescope (EHT). Het EHT-project exploiteert een wereldwijd zeer lange basislijn interferometrisch (VLBI in het engels) telescoop-netwerk op een golflengte van 1 mm dat de accretie van zwarte gaten en emissieprocessen bestudeert en uiteindelijk Einsteins theorie van de algemene relativiteitstheorie (GR in het engels) test via directe beeldvorming van de schaduwen van zwarte gaten.

In hoofdstuk 2 presenteren we de analyse van 3 mm-waarnemingen van Sgr A* met de Global Millimeter VLBI Array (GMVA) vergezeld door de zeer gevoelige Atacama Large Millimeter Array (ALMA) in 2017, complementair aan de EHT. Waarnemingen op 3 mm zijn ideaal om het al lang bestaande debat over de oorsprong van het primaire radio-emissieproces in Sgr A* toe te lichten, aangezien de beeldmorfolgie niet wordt gedomineerd door de schaduw van het zwart gat (zoals mogelijk het geval is op 1 mm) of gedomineerd door effecten van verstrooiing in het interstellaire plasma (zoals het geval is bij langere golflengten). Met de verbeterde resolutie en gevoeligheid verkregen met ALMA, hebben we voor het eerst een intrinsiek beeld van Sgr A* gereconstrueerd op 3 mm in de eerste publicatie op elke golflengte die ALMA gebruikt als een element van een VLBI-array. Mijn werk gebruikte nieuw ontwikkelde beeldvormingstechnieken zoals verstrooiing mitigatie en tweede moment regularisatie (Hoofdstuk 3) om de intrinsieke structuur van Sgr A* te bepalen. We hebben aangetoond dat het intrinsieke beeld en de onderliggende gegevens de mogelijke emissiemodellen voor Sgr A* limiteren tot accretieschijf- of bijna face-on jet-gedomineerde emissiemodellen.

In hoofdstuk 3 presenteren we de regularisatietechniek van het tweede beeldmoment, ontwikkeld om de beeldvorming van compacte bronnen met VLBI te ondersteunen. Voor onze GMVA+ALMA-waarnemingen hielp het bij de beeldvorming van Sgr A*-gegevens, waar korte-
basislijninformation werd gedomineerd door lage signaal-ruismetingen die de beelden naar niet-fysische intensiteitsverdelingen drijven. De techniek is ook een handig hulpmiddel voor het afbeelden van EHT-gegevens: de beeldgetrouwheid van de EHT wordt momenteel bepaald door zijn schaarse basislijndekking, en is zeer gevoelig voor atmosferische omstandigheden en verlies van locaties tussen experimenten. Ons algoritme dwingt een specifiek tweede moment af op het gereconstrueerd beeld in de vorm van een groottebeperking, en maakt het mogelijk om informatie te herstellen die verloren is gegaan in hiaten van de basislijndekking bij korte basislijnen. Deze regularisatietechniek op het tweede moment is een bijzonder krachtig hulpmiddel voor dynamische (video) reconstructies, waarbij individuele momentopnames een zeer beperkte dekking hebben. Onze methode en wordt momenteel gebruikt voor het verkrijgen van beelden van Sgr A* met de EHT.

In hoofdstuk 4 presenteren we studies over intrinsieke morfologie en interstellaire verstrooiing van Sgr A* met GMVA+ALMA-waarnemingen van 2017 en 2018 bij de golflengte van 3mm. Onze nieuwe waarnemingen uit 2018 bevestigen de niet-Gaussische structuur in het verstrooide beeld dat in 2017 werd gezien, wat een strikte beperking op de brongrootte en een bovengrens op de dissipatieschaal van interstellaire turbulentie biedt. Tegelijkertijd met de verstrooiingsparameters, vinden we op zijn hoogst een licht asymmetrische intrinsieke bronmorfologie voor Sgr A*. Naast de algemene beeldmorfologie, sluit het vergelijken van waarnemingen uit 2017 en 2018 met voorspellingen van verstrooiingsmodellen afdoende een model uit met sterke kleinschalige beeldvervormingen, dit betekent dat verstrooiing slechts een zwakke invloed zal hebben op 1mm-afbeeldingen met de EHT.

In april 2019 publiceerden we de eerste 1mm-beeldvormingsresultaten voor het SZG in het centrum van het M87-sterrenstelsel, waarbij een schaduw van een zwart gat werd onthuld die werd geproduceerd door extreme zwaartekrachtlenzen. In hoofdstuk 5 presenteren we de gegevenskalibratiemethoden die speciaal zijn ontwikkeld om de unieke eigenschappen van de EHT-waarnemingen aan te pakken, de ruggengraat van de 12 ordes van grootte in dataeductie tussen onbewerkte opnames en de inmiddels beroemde afbeelding van het M87* zwarte gat. Naast de moeilijkheid van campagnecoördinatie, acquisitie en enorme hoeveelheid gegevens, maken de heterogeniteit van de EHT-array en zijn gevoeligheid voor weersomstandigheden en atmosferische turbulentie bij 1mm de gegevenskalibratie bijzonder moeilijk. De correlatiefase, waar de onbewerkte telescoopgegevens worden gecombineerd en gemeenschappelijke signalen worden gedetecteerd, wordt uitgevoerd op twee centrale rekenfaciliteiten. Er zijn drie onafhankelijke pijplijnen ontwikkeld om instrumentele en atmosferische vertragingen in de aankomsttijd van de signalen bij de telescopen te corrigeren, voortbouwend op verouderde EHT, laagfrequente VLBI en nieuw beschikbare verwerkingsalgoritmen. Parallelle kalibratiepijplijnen maakten een uitgebreide reeks kruisvalidatietests mogelijk om de datakwaliteit en systematiek het best te kwantificeren. Om het intrinsieke bronsignaal van het telescoopgedrag te ontwarren, werden gedetailleerde studies uitgevoerd naar telescoopoperaties, gevoeligheden en observatieomstandigheden. De uiteindelijke gekalibreerde M87*-gegevens, klaar voor wetenschappelijke analyse en beeldvorming, vertonen duidelijke indicaties van een asymmetrische ringachtige structuur, met kleine structurele variaties in de loop van de waarnemingscampagne.

In hoofdstuk 6 presenteren we een subset van de analyse- en validatieprocedures voor de totale intensiteit en polarimetrische beeldvorming van het M87* zwarte gat, waarbij we ons con-
Samenvatting

centren op gegevens en instrumenteigenschappen als onderdeel van het beeldvormingsproces. Onze totale intensiteitsbeelden voor 2019 tonen een prominente ring met een diameter van ~40 microboogseconden met een verbeterde helderheid in het zuidelijke deel, aanhoudend gedurende vier observatiemachten. Om de betrouwbaarheid van deze resultaten te beoordelen, hebben we een beeldvormingsprocedure in twee fasen geïmplementeerd. In de eerste fase produceerden vier teams onafhankelijke beelden van M87* met behulp van zowel de gevestigde CLEAN-beeldvormingsmethode als nieuwere regularized maximum likelihood (RML) methoden. Deze fase vermeed menselijke vooringenomenheid door de communicatie tussen teams te beperken en stelde ons in staat gemeenschappelijke kenmerken van onafhankelijke reconstructies te beoordelen. In de tweede fase hebben we synthetische gegevens gereconstrueerd waarbij het onderliggende model bekend was uit een groot aantal van beeldvormingsparameters te onderzoeken. We hebben parameters objectief geselecteerd die synthetische beelden opleverden die dichtbij het onderliggende model lagen. Deze parameters hebben we dan gebruikt bij het reconstrueren van afbeeldingen van M87*. De gereconstrueerde stationsgedragingen verkregen uit de uiteindelijke beelden werden vergeleken met gedragingen gereconstrueerd uit kalibratiequasar 3C 279, en vertoonden grote consistentie zoals verwacht. Onze polarimetrische afbeeldingen van M87* uit 2021 toonden een spiraalpatroon van de positiehoeken van de elektrische vector en een lineaire piekopolarisatie van 15% geconcentreerd in het zuidwestelijke deel van de ring. Polarimetrische beeldvorming van het M87* zwarte gat volgde een soortgelijk tweestap validatieproces: de voorlopige resultaten werden verkregen door opsplitsing in drie algoritmespecifieke teams, waarbij de nadruk lag op overeenstemming over afgeleide instrumentele lekkages van de EHT-stations; vervolgens werd een synthetische data-scan uitgevoerd om optimale parameters te selecteren. De gereconstrueerde stationlekkages verkregen uit de uiteindelijke afbeeldingen werden vergeleken met die afgeleid uit beeldvorming van kalibratoren als onderdeel van onze uitgebreide validatietesten.

Dit is slechts het begin voor de EHT-wetenschap. We bereiden momenteel de publicaties voor van resultaten op horizon-schaal van Sgr A*, die, in combinatie met onze belangrijke partners op andere golfstanden, inclusief de 3 mm-inspanning, enkele inzichten zullen verschaffen in zeer opwindende wetenschap. De EHT zal de komende jaren blijven observeren en uitbreiden in het aantal telescopen en door 0,8 mm waarneming, met een volgende generatie EHT-project in de plannings- en ontwerpfase voor de komende twee decennia. Met deze technische uitbreidingen hopen we films te maken van onze twee hoofddoelen en te verbeteren voor het observeren van de schaduw van het zwarte gat voor GR-tests en van de uitgebreide structuur voor jet- en accretiestudies mogelijk te maken.
I was born in Boghni, Algeria, on 27 May 1994. In 2001, at the age of seven, my family immigrated to Montreal, Canada, where I continued my schooling. At the age of eight, I developed a keen interest for the Solar System from a school project, which fed a passion for astronomy that persists until today. At the age of twelve, my father gifted me a telescope in support of my passion, which now remains one of my most precious possessions.

I went to secondary school at the Montreal International School until 2008, when my family immigrated to the Netherlands. I finished my secondary schooling at the Arnhem International School (now Rivers International School), where I graduated with an International Baccalaureate focused on physical sciences and mathematics. To pursue a career in astronomy, I studied at McGill University in Montreal and earned a Bachelor’s degree in physics (with a minor in economics that proved very useful) in 2015. In my physics curriculum, we did not have any astronomy courses, and therefore I lacked real research experience in astronomy. In 2014, I wanted to do an astronomy project as an undergraduate student while I was back in the Netherlands for summer vacation, and got in touch with Prof. Heino Falcke at Radboud University. Under his supervision, I learned about radio astronomy and the Event Horizon Telescope (EHT) project. This led me to apply to Radboud for a master’s study in astrophysics in 2015, where I was accepted with a scholarship. In 2017, I completed my Master’s degree in astrophysics in Prof. Falcke’s group, where I was mentored by Dr. Ciriaco Goddi on interferometric data calibration and reduction. During my master’s studies I got involved in EHT observations and telescope operations, and became a point of contact for the Submillimeter Telescope (SMT) in Arizona. My first observations with the EHT were in 2016, where I traveled to Arizona to work on telescope calibration and help operations during an engineering observation to bring the ALMA telescope into the array. I then observed for every science campaign since then as part of the SMT EHT staff, and I am now deputy lead for SMT operations within the...
EHT efforts. I stayed on with Prof. Falcke as a PhD student to work on EHT data science and help make the first images of black holes.

I had many opportunities to travel and interact with members of the collaboration and the wider scientific community. In 2018, I visited the Smithsonian Astrophysical Observatory in Boston as a Predoctoral Fellow, under the supervision of Dr. Michael Johnson. I stayed in Boston for seven months in 2018, and I was fortunate to be at one of the epicenters of EHT activities: I was able to take leadership roles in the data calibration and imaging efforts, as well as make tremendous progress on my own personal science projects. This was a period of rapid growth for me, both personally and professionally. In April 2019, when the nerve-wracking moment of revealing the first image of a black hole came, it became a turning point for my career. Having played a central role in the story of the EHT’s now-famous results, I am so grateful to have gotten amazing opportunities to give talks at major colloquia, conferences, and public events around the world. It was also an extremely enriching learning experience that improved my communication skills, my understanding of the importance of outreach, and our role (and responsibility) as scientists to give back to the community at large. Shifting gears toward the M87 polarization results was another growth opportunity, where I again played a major role. This year, presenting the first polarized images of a black hole, a challenging feat, put me center-stage alongside my colleagues. I am very honored and grateful for the many opportunities brought by EHT science to grow as a scientist, to be visible in the public eye, and to be part of this incredibly unique adventure.

Outside of scientific research, I have performed a number of academic service duties. I have led the organization of several EHT workshops and conferences, for a majority of which I was the only graduate student. I have presented collaboration work at multiple topic-specific internal reviews and assessments over the past three years, and I was the only graduate student to serve on the EHT Project Director search committee. During my PhD I also taught a number of physics and astronomy courses as a teaching assistant, a very fulfilling and enjoyable experience. I have been also very committed to astronomy outreach projects, actively interacting online and offline with the general public in the form of public talks, camp lectures for young girls, informative Twitter threads, and educational audiovisual material to understand our EHT science. With the EHT, we have the immense advantage that we hold a powerful source of general interest: our images of black holes continuously yield a lot of engagement with our material.

Now that my PhD studies at Radboud University are completed, I will be returning to Boston, at the Center for Astrophysics | Harvard & Smithsonian, where I will continue my research as a NASA Einstein Fellow. Until now, my individual science had been put on the back-burner for the good of EHT scientific progress, but as an Einstein Fellow I will have the independence to pursue other avenues and expand my scientific portfolio. I am very much looking forward to this new chapter of my career.
LIST OF PUBLICATIONS

Primary Author Publications


Collaboration Publications


List of Publications


Other Publications


List of Publications


279
Technical Documentation


I walked and somehow managed to reach the sea. The journey through my scholarly studies is coming to an end, and a new one will begin. It feels a bit surreal to finally reach this end point that seemed so far ahead all this time. I did not walk alone, alongside me, supporting me, were countless people, some from the very beginning, some I’ve had the pleasure of meeting along the way. I’d like to take this space to thank them for making me the person I am today.

First, I thank my PhD supervisor Heino Falcke for his support and guidance both in my research and in my professional goals. I met Heino by coincidence (or he would say ‘fate’) as I was looking for an astronomy project in 2014, when I was still an undergraduate student. Heino has made me the independent scientist I am today: his complete trust and support, even when disagreements occurred (usually concluding in my victory), has allowed me to branch out and stand on my own two feet. Over the years, Heino has provided so many opportunities to make my place and grow within the EHT collaboration, to collaborate with top-tier scientists, and to take a leading role in science communication on the world stage. I’ve enjoyed a wonderful working relationship with Heino that is very different from the typical PhD student-advisor dynamic: I’ve always felt listened to and treated with great respect as a scientific equal. Our discussions were always very enriching, but in the end the decisions always came down to what I felt was best. Without his unfaltering support I would not have been able to pack the almost 300 pages of work this book contains.

I also thank Michael Johnson for his incredible mentorship and support throughout my PhD studies. I met Michael at the 2016 EHT collaboration meeting, where he told me he had read my telescope calibration memo and called it a “very fun read”. After spending 3 years working on scattering with Michael, I now understand how that memo can be considered light reading. Michael taught me everything I know about imaging techniques, Sgr A* science and interstellar scattering. He helped me build my scientific and technical foundation and guided me through the workings of the EHT collaboration. Michael, I had the privilege to work very closely with you for the first time in 2018 when you hosted me at the CfA, and we’ve had such a friendly working relationship ever since, thank you so much for everything. I am really thrilled to be going back to the CfA and continue this adventure.

To my Radboud/BlackHoleCam community, thank you all for your support and friendship and abundant cake. Monika, thank you for being such a great role model for an early-career woman like me. I am so happy that I had the opportunity to interact with you, and work side
Acknowledgments

by side on the polarization results and teaching. Not only that, but also seeing your strength, brilliance, and leadership inspired me to grow in many ways. I thank my PhD brothers, Freek (the guitarist from Blueshift), Michael (King Janssen), and Jordy (our meme dealer) for their support, humor, company and friendship throughout our common studies. You all started your PhDs one year before me, but I always felt like I was going through all this alongside you, as we grew and paved our way inside the EHT together. Christiaan and Thomas, thank you for your enriching conversations, always taking time to explain things that went right over my head, and strange dream recounts. Raquel, thank you for teaching me about interferometry back in 2014 when I jumped into this project, and for your friendship and common tastes. Thank you Remo and Ciriaco for mentoring me in telescope and data calibration. André, Daan, Cornelia, Gijs, Shan-Shan, Hector, Alejandra and Noemi, thank you for making our local group so vibrant and diverse and supportive. Esther, Katharina, and Amanda, without you I would not have smoothly finished my PhD, thank you for always looking out for me and making sure I’m on the right track. To Gijs, Marja, Monique, and Merijn, thank you for your support to the department, and especially for all the efforts you put in to make us feel connected this past year, as we all went through very challenging times. To my office mates over the years, Payaswini, Andrei, Irene, Michael, Freek, Anastasia, and Anne, thank you for the positive atmosphere, jokes, plant jungle, occasional karaoke, and friendliness that made our office such an enjoyable place of work. I have missed it very much this past year.

To my SAO/CfA community, thank you all for your support and friendship and late night collective hallucinations.1 Lindy, thank you for your support and mentorship on the observational side. Thank you for always lending an ear whenever I needed advice, and for always making me laugh with your dry jokes and absurd conversation segues. I am really proud of the tremendous feats we accomplished with just the five of us, you, Maciek, Michael, CK and me. Maciek, thank you for your friendship, strange crocodile songs, great guitar covers, and support. Kazu, thank you for taking me under your wing, I enjoyed collaborating with you on the AIPS pipeline, leading our imaging team, and designing our super cool team merch that obviously made Team 2 the best. Your positivity and smiley (SMILI?) attitude is always a welcome energy boost. Katie and Andrew, thank you for your friendship, late night singing sessions, and teaching me so much about imaging. It was a real pleasure to work with you both and I’m looking forward to interacting more in the near future. Daniel, Dom, and Joseph, thank you for your friendship and jokes and snacks (either Hmart goodies, or oreos, or cheerios) and bonding over movie nights. Shep, thank you for being so welcoming at SAO, for your support and contagious enthusiasm, and for the great opportunities you have provided me to grow as a scientist. I am really excited to return to SAO and be part of the vibrant team you have created.

It is impossible to thank everyone I’ve collaborated with by name, there are just so many. I’d like to thank my wonderful collaborators, mentors and colleagues in the EHT collaboration. I’ve had the chance to work with so many talented and friendly people, and I am really happy to have made some great personal and professional connections among them. Dan, thank you for always being so welcoming in Arizona, for mentoring me in telescope operations and EHT observing, including learning to keep you awake while you drive up a mountain, and being someone I can

1Shout out to Scatter J, Sparse K, Katie Snowman, MEM87, Lindyhops and Maciek Seaborn, from Sarapedia.
always talk to for support and advice whenever I needed. Junhan, thank you for your friendship and company during long observing schedules at SMT, we’ve had lots of fun at the telescope and I miss the mountain greatly. To all the wonderful women in the collaboration, in particular Monika, Katie, Sera, Feryal, Jess, Svetlana, Raquel, Kari, Fumie, Ilse, Silke, Izumi, Lia, Kazi, and Alejandra, with whom I’ve had the pleasure of interacting many times on various fronts, thank you for being a daily inspiration and support system, for your kindness and listening ear, and for continuing to break stereotypes of what scientists are and do everyday. I am proud of working in this project alongside you all.

To my dear friends, who have supported me throughout my studies and dealt with so many of my complaints in loud venting sessions, thank you for your support and friendship. Thank you especially to Domenique, Nerissa, Abigail, Caroline, and Daniel for always finding time to reconnect since high school, even if our lives have taken us in so many different paths. Ben, thank you for being a source of laughter and camaraderie as we struggled through astrophysics grad school together. Thank you to my ARMY friends who made the pandemic this last year feel so much less lonely, through book clubs, streaming parties, birthday celebrations, spa nights, secret Santa, losing all our money from collecting, and many fun late night chats. In particular, my KG girls, AWO friends, and PMC family, I appreciate your friendship and support so much. To BTS, who have been with me since undergrad and helped me through tough times, thank you for always inspiring me to be better and do better, you showed me I have reasons to love myself.

Finally, to my loving family, Mom, Dad, Melissa, thank you for always supporting me through this roller-coaster that was my time as a PhD student. You’ve always been amused by my tenacity to become an astronomer since I was little, and I am so grateful to you for allowing me to chase my dreams and supporting me through all the decisions and sacrifices to get me there. Without you, none of this would have been possible, I love you to M87 and back.