

The Nature and Impact of Active Galactic Nuclei

by

Paul Tiede

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics and Astronomy

Waterloo, Ontario, Canada, 2021

© Paul Tiede 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Anthony Readhead
Robinson Professor of Astronomy,
The Division of Physics, Mathematics and Astronomy,
California Institute of Technology

Supervisor: Avery Broderick
Associate Professor, Dept. of Physics and Astronomy,
University of Waterloo;
John Archibald Wheeler Chair of Theoretical Physics,
Perimeter Institute for Theoretical Physics

Internal Member: Niayesh Afshordi
Associate Professor, Dept. of Physics and Astronomy,
University of Waterloo;
Associate Faculty, Perimeter Institute for Theoretical Physics

Internal Member: Brian McNamara
Professor, Dept. of Physics and Astronomy,
University of Waterloo

Internal-External Member: Francis Poulin
Professor, Dept. of Applied Mathematics,
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This dissertation is the partially the product of 2 co-authored publications and 3 papers in preparation for publishing:

- Chapter 2 is based on Tiede, P., Broderick, A. E., & Palumbo, D. C. M. 2020, arXiv:2012.07889, Variational Image Feature Extraction for the EHT which is currently going through peer review at the Astrophysical Journal
- Chapter 3 is based on Tiede, P., Broderick, A. E., Palumbo, D. C. M., & Chael, A. in prep.a, Measuring the Ellipticity of the M 87 EHT Images. A paper currently in preparation
- Chapter 4 is based on Tiede, P., Broderick, A. E., & Saifuddin, S. in prep.b, Next Generation Samplers for THEMIS. A paper that is currently in preparation
- Chapter 5 is based on Tiede, P., & Broderick, A. E. in prep., Bayesian Ellipticity Estimates of M 87. A paper that is currently in preparation
- Chapter 6 is based on Tiede, P., Pu, H.-Y., Broderick, A. E., et al. 2020, The Astrophysical Journal, 892, 132. <https://doi.org/10.3847/1538-4357/ab744c>

For all publications, I was the primary author, and was responsible for the majority of the analysis, results, and writing. My co-authors provided valuable discussion, and help with editing.

Abstract

The gravitational interaction around the event horizon of black holes presents theoretical challenges. With the advent of the Event Horizon Telescope (EHT), we are now entering an era in physics where we can probe the structure of spacetime on horizon scales. The EHT presents the first opportunity to directly image the supermassive black holes at the center of the Milky Way and M 87. By imaging the central black hole, we can directly learn about the nature of spacetime and plasma physics on horizon scales.

The black hole images produced by the EHT are dominated by a bright ring. The ellipticity of the ring could potentially signal deviations from general relativity. However, whether the EHT imaging techniques can robustly detect ellipticity has not been fully explored. Chapters 2–5 analyze the EHT’s ability to measure ellipticity in four parts. First, in [Chapter 2](#), we develop a method to extract image features (e.g., ring ellipticity) called *variational image domain analysis*. Second, in [Chapter 3](#), we apply variational image domain analysis to the M 87 image reconstruction pipeline and demonstrate that it is unable to measure ellipticity. The core reason for this failure is that traditional radio imaging techniques cannot quantify image uncertainty. To solve this issue, in Chapters 4 and 5 we use THEMIS, a Bayesian parameter estimation framework for the EHT, to robustly measure the ellipticity of M 87. To apply THEMIS to the problem of Bayesian imaging, we developed a new sampler interface in [Chapter 4](#). In [Chapter 5](#) we apply THEMIS to M 87 and construct the first Bayesian estimates of its ellipticity. Furthermore, we demonstrate that the measured ellipticity is consistent with the expected ellipticity from an accretion disk around a Kerr black hole.

In [Chapter 6](#) we describe a novel method to measure spacetime around Sgr A* using hot spots. While M 87 is static over an observation, Sgr A* is dynamic, changing on minute timescales. Furthermore, Sgr A* flares 1–3 times a day in sub-mm, infrared, and X-ray. The Gravity Collaboration recently demonstrated that hot spots near the innermost stable circular orbit explain Sgr A* flares. Using THEMIS, we construct an efficient semi-analytical model of hotspots and fit simulated Sgr A* data from the 2017 EHT observations. We demonstrate that the EHT could potentially make a sub-percent spin measurement of Sgr A* by tracking the evolution of these flares. Furthermore, by observing multiple flares, we can tomographically map spacetime around Sgr A*, providing a test of general relativity in the strong-field regime.

Acknowledgements

This thesis would not have been possible without the support from many people. In particular, I would like to thank my supervisor Avery Broderick, who got me interested in high-energy astrophysics when I was an undergraduate student. Their guidance during my academic career has been invaluable, and my journey in academia would not have been possible without it.

To my fellow Waterloo and Perimeter colleagues Alex Preciado, Boris Georgiev, Britton Jeter, Chunchong (Rufus) Ni, Hung-Yi Pu, and Roman Gold, I would like to thank you for the immense work and tireless effort during the analysis of the M 87 data. I would also like to thank Mansour Karami, Dom Pesce, and Saifuddin Syed for getting me interested in Markov Chain Monte Carlo and its applications to EHT data. Additionally, I would like to thank the entire Event Horizon Telescope Collaboration and the academic and professional growth it provided during my Ph.D.

Alexandra White, Niki, and Nala, your support, comfort, and patience during my Ph.D. gave me the energy to push through and finish during a global pandemic.

Dedication

Dedicated to my mother, father, brother, oma and opa, who have supported my academic journey, without them this would not have been possible.

Table of Contents

List of Figures	xii
List of Tables	xiv
1 Introduction	1
1.1 Gravitational Physics	1
1.2 Active Galactic Nuclei	3
1.3 The Event Horizon Telescope	5
1.4 THEMIS: An EHT Bayesian Parameter Estimation Framework	8
1.4.1 Bayesian Inference	8
1.4.2 Markov Chain Monte Carlo	9
1.4.3 THEMIS	11
1.5 Summary	13
2 Variational Image Domain Analysis for the Event Horizon Telescope	15
2.1 Introduction	15
2.2 Variational Image Domain Analysis	17
2.2.1 Image Templates in VIDA.jl	18
2.2.2 Probability Divergences	22
2.2.3 Optimizing the Divergence	23
2.2.4 REx and relating to VIDA parameters	23

2.3	Validating VIDA	26
2.3.1	Step 1: Selecting Ground Truth Images	26
2.3.2	Step 2: Creating Simulated EHT Observations	27
2.3.3	Step 3: Generating an Ensemble of Reconstructed Images from Simulated VLBI Data	27
2.3.4	Step 4: Applying VIDA to the Ring-Like Image Reconstruction Ensembles	30
2.4	Applying VIDA to Additional Simulated Image Reconstructions	31
2.4.1	Double Gaussian	31
2.4.2	Disk Image	33
2.5	Summary and Conclusions	36
3	Constraining the ellipticity of M 87 from EHT Image Reconstructions	37
3.1	Introduction	37
3.2	Background	38
3.2.1	Image Reconstructions and the M 87 top set	39
3.2.2	Feature Extraction Techniques	40
3.2.3	Review of M 87 Ellipticity Measurement	40
3.3	Geometric Test	44
3.3.1	Asymmetric Image Test	44
3.3.2	Geometric Results	44
3.4	Calibrating the M 87 Ellipticity Measurement	45
3.4.1	Scaled set	45
3.4.2	Stretched set	45
3.4.3	Removing failed top set reconstructions	46
3.4.4	Scaled set results	48
3.4.5	Stretched set results	50
3.5	Summary and Conclusions	52

4	Next Generation Samplers for Themis	54
4.1	Introduction	54
4.2	Abstract Sampler Framework	55
4.3	Local Samplers	56
4.3.1	Hamiltonian Monte Carlo	56
4.3.2	Automated Factor Slice Sampler	59
4.4	Global Samplers: Parallel Tempering	62
4.4.1	Review of Parallel Tempering	62
4.4.2	Optimizing Parallel Tempering	64
4.4.3	Implementation in THEMIS	67
4.4.4	Comparison to THEMIS' Previous PT Sampler	67
4.5	Assessing MCMC Convergence	68
4.6	Bayesian Imaging Validation	69
4.7	Future Improvements	73
4.8	Conclusion	74
5	A Bayesian Image Estimate of M 87 ellipticity	76
5.1	Introduction	76
5.2	Bayesian Imaging and Feature Extraction	77
5.2.1	Image Domain Model	77
5.2.2	Image Feature Extraction	79
5.3	Data	79
5.4	Geometric Tests	81
5.4.1	Pixel Optimization	81
5.4.2	Orientation Dependence of Bayesian Imaging	85
5.5	M 87 Imaging	85
5.5.1	Interpreting the M 87 ellipticity	88
5.6	Conclusion	90

6 Space Time Tomography with the Event Horizon Telescope	91
6.1 Introduction	91
6.2 Hot spot model	93
6.2.1 Density profile evolution	93
6.2.2 Radiative transfer and Ray-tracing	96
6.3 Shearing hot spots with the EHT	99
6.3.1 Creating Synthetic EHT Data	100
6.3.2 Extracting spacetime and spot parameters with THEMIS	101
6.3.3 Adding Systematics – Scattering and Background flows	102
6.3.4 The impact of disk inclination	109
6.4 Spacetime tomography	113
6.4.1 Constructing a synthetic tomographical map of spacetime	114
6.4.2 Results	116
6.5 Conclusions	117
7 Conclusion	119
References	121
Appendices	134
A Complex Gain Marginalization	135
B Bayesian Information Criterion	137

List of Figures

1.1	World map with the 2017 EHT array stations highlighted.	6
1.2	Layout of THEMIS.	12
2.1	Example VIDA run	19
2.2	Comparing REx and VIDA ellipticity measurements	24
2.3	VIDA validation images	28
2.4	VIDA and REx validation results for ring images	29
2.5	VIDA non-ring validation images	32
2.6	VIDA results for double gaussian validation image	33
2.7	VIDA results for disk validation image	34
2.8	VIDA applied to average disk image	35
3.1	EHT coverage of M 87 on April 11 with coverage gaps circled	39
3.2	<code>eht-imaging</code> reconstructions of M 87 with ellipticity measurements	41
3.3	Elliptical ring reconstructions using <code>eht-imaging</code> pipeline	42
3.4	CNN based image classification of <code>eht-imaging</code> reconstructions	46
3.5	Impact of reconstruction classification on ellipticity	47
3.6	Joint probability of ellipticity and its orientation for scaled GRMHD reconstructions	49
3.7	Residual distribution of the stretched GRMHD top sets	51
3.8	Intrinsic ellipticity τ vs recovered ellipticity of the stretched GRMHD images.	52

4.1	Slice sampler visualization	60
4.2	Tempering demonstration	63
4.3	Example DEO diagnostic plots	70
4.4	Sampler Bayesian imaging benchmark	71
4.5	MCMC benchmark of a Bayesian imaging of M 87	72
5.1	Comparison of Bayesian reconstruction of elliptical rings for different numbers of raster pixels	82
5.2	Comparison between <code>eht-imaging</code> and THEMIS' reconstructions of the elliptical ring simulated data sets	84
5.3	April 11 LO band example Bayesian image reconstructions of M 87	86
5.4	Ellipticity and orientation of M 87 from Bayesian imaging	87
5.5	Bayesian ellipticity measurements of M 87	88
5.6	Comparison of M 87 ellipticity to GRMHD simulations	89
6.1	Example frames from a shearing hot spot movie	93
6.2	VLBI observables of a hot spot movie	99
6.3	Joint probability distribution for all parameters from a simulated hot spot	103
6.4	Hot spot movie frame and their visibility amplitudes	104
6.5	Impact of scattering and RIAF on observables	106
6.6	VLBI lightcurves for a shearing spot with blurring and a RIAF background	107
6.7	The impact of Sgr A* systematics on hot spot measurements	108
6.8	Hot spot simulations observed at different inclinations	109
6.9	Recovered parameter distribution of hot spot observed at different inclinations	110
6.10	Hot spot movie with refractive scattering included	111
6.11	Impact of varying spin and initial radius on recovered hot spot parameters	113
6.12	Posteriors for simulated hot spots around a spin 0, 0.5, 0.9 black hole . . .	115
6.13	Example spacetime tomography map	117

List of Tables

1.1	EHT 2017 Telescopes	6
3.1	Recovered elliptical ring parameters for ellipticity validation test.	43
5.1	Elliptical Ring Raster Survey Results	83
6.1	Spacetime & Hot Spot Parameters	112

Chapter 1

Introduction

1.1 Gravitational Physics

Modern physics is described by two pillars: quantum field theory and general relativity (GR). However, after more than half a century of research (Rosenfeld, 1930; Heisenberg, 1938; Bergmann & Brunings, 1949; Dirac, 1950; Bergmann & Goldberg, 1955), general relativity has not been unified with quantum field theory. From a theoretical perspective, the challenges are both conceptual and technical. Technically, quantum gravity requires advanced mathematical techniques such as topological quantum field theories (Witten, 1988). Conceptually, how to describe time within a quantum theory of gravity it still unclear. In addition, while there are theoretical motivations for the failings of GR (e.g., non-renormalizability Hooft, 1973; Hooft & Veltman, 1974), there are no conclusive observations that contradict its predictions. For example, cosmological expansion and structure formation can be explained by a cosmological constant and dark matter. Furthermore, while there is tension between the local and early universe estimates of the Hubble constant (Di Valentino et al., 2021), this measurement is still uncertain.

While GR has been continuously tested on solar system and cosmological scales, only recently has it been probed on horizon scales. With the advent of LIGO, and the Event Horizon Telescope (EHT), probing physics on horizon scales is now possible. Theoretically, the event horizon of black holes has lead to many interesting hypotheses, such as Hawking Radiation (Hawking, 1976), the no-hair theorem (Israel, 1967), and the information paradox (see Mathur, 2009, for a review). With the first detection of gravitational waves (Abbott et al., 2016), tentative evidence suggests that novel physics may be occurring on horizon scales (Abedi et al., 2017). Any deviations of GR on horizon scales would provide

the first experimental evidence for a regime where GR is not valid. In this thesis, we will explore what measurements the EHT can make on horizon scales.

All astrophysical black holes are described by the Kerr metric (Kerr, 1963):

$$g = - \left(1 - \frac{2Mr}{\rho^2} \right) dt^2 - \frac{4Mar \sin^2 \theta}{\rho^2} dt d\phi + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + \frac{\Sigma}{\rho^2} \sin^2 \theta d\phi^2, \quad (1.1)$$

where $\rho = r^2 + a^2 \cos \theta^2$, $\Delta = r^2 - 2Mr + a^2$, and $\Sigma = (r^2 + a^2)^2 - a^2 \Delta \sin^2 \theta$. This metric is parameterized by two numbers: mass M and spin a . This implies that to external observers, black holes show no evidence of their formation. This lack of information is related to the information loss paradox (Hawking, 1976). At the core of this paradox, is the nature of the event horizon and the applicability of GR on horizon scales (Mathur, 2009; Almheiri et al., 2013).

In addition to the event horizon, key qualitative features of the Kerr metric include the ergosphere, photon sphere, and the innermost stable circular orbit (ISCO). The ergosphere refers to the region around the black hole where all material must rotate in the same direction as the black hole spin. The ISCO is the innermost radius in spacetime where massive particles can move on stable circular orbits. The photon sphere is a compact 2D surface in Kerr spacetime where photons travel on closed orbits around the black hole. For a non-spinning black hole, all photon orbits occur at a Schwarzschild radius of $r = 3M$. For spinning black holes the photon sphere flattens into an ellipsoidal surface. As a result, there exists a family of bound orbits characterized by an orbital radius, r , which satisfies:

$$r_-^\gamma \leq r \leq r_+^\gamma \quad (1.2)$$

$$r_\pm^\gamma = 2M \left[1 + \cos \left(\frac{2}{3} \arccos \left(\pm \frac{a}{M} \right) \right) \right]. \quad (1.3)$$

The existence of such a surface gives a black hole image its “shadow”. For an observer, this sphere appears as multiple lensed rings (Luminet, 1979; Teo, 2003). Following the notation of Johnson et al. (2020), these rings are labeled by the number of half-orbits the light traverses around the black hole. For example, the $n = 1$ ring corresponds to photons that do a single half orbit, while $n = 2$ is a full orbit. As $n \rightarrow \infty$, this converges to the photons that live on the photon sphere. For a non-spinning black hole, this visually appears as a infinitely thin ring at a radius of $r = \sqrt{27}M$. The shape and relative spacing between these nested rings can uniquely measure the spin and mass of the central black hole (Johnson et al., 2020; Broderick et al., 2021). Furthermore, deviations from GR can alter this relation, producing additional ellipticity or alter the spacing of the rings (Johannsen &

Psaltis, 2011; Johannsen, 2013; Psaltis et al., 2015; Medeiros et al., 2020; Broderick et al., 2021).

Another way to measure the impact of this photon sphere is through time delays. Light emitted from an object around the black hole will orbit on different trajectories depending on its initial direction. As a result, to an external observer, the light from the object splits into multiple images seen at different times. This effect leads to a strong-gravity version of the Shapiro delay and is known as black hole glimmer (Moriyama et al., 2019; Wong, 2021). Glimmer is most visually apparent when, e.g., a localized flaring event occurs around a black hole. The different arrival times and locations of the flare emission around the black hole, uniquely characterize the black hole’s mass, spin, and inclination. However, measuring glimmer and the black hole shadow require resolving the central black hole’s horizon.

Interestingly, the shadow of the central black hole is related to the quasi-normal modes or ring-down spectrum. This result was originally shown for Schwarzschild spacetimes in Stefanov et al. (2010), and more recently for Kerr spacetime (Yang, 2021). This result demonstrates that while LIGO and the EHT use different observations, both probe the same gravitational observables. However, while LIGO is interested in solar mass black holes, most of the known black holes live at the centers of galaxies and have masses exceeding $10^6 M_\odot$.

1.2 Active Galactic Nuclei

Active galactic nuclei (AGN) are some of the most consistently luminous objects in the night sky. Seyfert galaxies (Seyfert, 1943) are a class of AGN, where the central object has a total intensity similar to the entire galaxy. Quasars, another form of AGN, first appeared as star-like objects with extremely high luminosity and varied on timescales less than a day. Soon after their discovery, it was realized that quasars had an extragalactic origin (Schmidt, 1963).

The standard model of AGN asserts that at their center is a supermassive black hole (Rees, 1984). The diverse behavior seen in AGN can hypothetically be explained by a small number of parameters. These parameters are related to the central black hole’s mass, spin, and the surrounding accretion environment. In Heckman & Best (2014) AGN are classified into two main groups: radiative and jet mode AGN. It is important to note that there are multiple subgroups within each group (e.g., radio-loud AGN and the existence of broad-emission line regions).

Radiative-mode AGN are classified by their relatively high luminosity $L/L_{\text{Edd}} \gtrsim 0.01$. It is thought that the dominant energy output from these sources is from a geometrically-thin and optically-thick accretion disk. In historical terminology, radiative-mode AGN would be classified as Seyfert's and quasistellar objects (QSO's). These AGN often have the “big blue bump”, a spectral feature characteristic of geometrically-thin accretion disks. Additionally, the majority of these sources do not display strong jet-like emission, except at the highest luminosity (Heckman & Best, 2014).

Compared to radiative-mode AGN, jet-mode AGN have lower accretion rates and subsequently lower luminosity $L/L_{\text{Edd}} \lesssim 0.01$. A key feature of these lower accretion rates is that the protons are decoupled from the electrons and cannot radiate away most of their energy. As a result, the accretion disk heats up, becomes vertically thick, and the energy is advected towards the black hole (see Yuan & Narayan, 2014, for a review). This accretion state is known as a radiatively inefficient accretion flow (RIAF). RIAFs can generically support outflow, i.e., winds and jets (Blandford & Begelman, 1999; Yuan et al., 2002). This property gives jet-mode AGN their name.

Jet-mode AGN display a variety of emission properties related to the observer’s line of sight and accretion rate. For example, blazars are AGN where the observer’s line of sight is parallel to the black hole’s jet. Blazars are highly variable and display super-luminal motion in radio. Furthermore, they can also produce relatively large X-ray and gamma-ray fluxes. On the opposite end of the spectrum, are low-luminosity AGN (LLAGN).

The two primary targets for the EHT, M 87 and Sgr A*, are well known examples of LLAGN. For instance, given M 87’s enormous mass, $6.5 \times 10^9 M_{\odot}$, its core bolometric luminosity, $3 \times 10^{42} \text{ ergs}^{-1}$ (Prieto et al., 2016), is only $3.6 \times 10^{-6} L_{\text{Edd}}$. Similarly, by observing the motion of O and B stars near the Galactic center, Schödel et al. (2002); Ghez et al. (2003); Ghez et al. (2008) were able to strongly suggest that Sgr A* is a $4.1 \times 10^6 M_{\odot}$ black hole. Given this mass, the bolometric luminosity for Sgr A* is very low, $L \sim 10^{36} \text{ ergs}^{-1} \sim 10^{-9} L_{\text{Edd}}$. RIAFs (with non-thermal electrons) are able to explain the observed properties of Sgr A* (Narayan et al., 1998; Yuan et al., 2004; Broderick & Loeb, 2009; Broderick et al., 2011), and M 87 (Reynolds et al., 1996), and a variety of other LLAGN (e.g., Di Matteo et al., 2000; Di Matteo et al., 2001; Ho et al., 2003).

While M 87 and Sgr A* are both LLAGN, they do have several differences. M 87 is a giant elliptical galaxy, while the Milky Way is a spiral galaxy. Sgr A* is also a thousand times less massive than M 87. For the EHT, this implies that while M 87 is approximately static over a night, Sgr A* will constantly evolve. Additionally, the local environment around M 87 is very different. M 87 has a large number of neighboring globular clusters when compared to Sgr A*. Another distinction of particular interest is the existence of a

jet. While M 87 has a radio-loud jet (Curtis, 1918), Sgr A* has no visible jet emission. Given that both black holes are likely RIAFs, the cause of this difference is currently unknown. One explanation is that jets could be powered by black hole spin through the Blandford-Znajeck process (Blandford & Znajek, 1977). Therefore, a measurement of spin for Sgr A* and M 87 could have important implications for the formation of jets.

Sgr A*, while displaying constant stochastic variability, also has brief high-excitation periods, i.e., flares. Strong broadband flaring activity is observed 1–3 times a day, from sub-mm to X-ray (Marrone, 2006; Gillessen et al., 2006; Witzel et al., 2012; Witzel et al., 2018; Fazio et al., 2018). Broderick & Loeb (2005, 2006) suggested that these flares could be due to magnetic reconnection events in the accretion disk. Given the highly magnetized and the hot accretion disk, reconnection is expected and seen in high-resolution plasma simulations (Ripperda et al., 2020). These magnetic reconnection events would then locally accelerate electrons in the accretion disk, creating non-thermal “hot spots”. The hot spot model predicts that the center of light of Sgr A* should appear to wobble during a flare (Broderick & Loeb, 2005, 2006). Over a decade after its prediction, Gravity Collaboration et al. (2018) observed precisely this motion, suggesting that at least a subset of these flares could be explained through orbiting hot spots. One area that is still uncertain is the nature of these flares in the sub-mm. Namely, while the X-ray and infrared observations of flares are known to be correlated, their relation in the sub-mm is uncertain (Eckart et al., 2008a,b).

The unknown questions raised above, namely the black holes spin and the nature of sub-mm flares, require observing the black hole on horizon scales. While spin measurements have been made for supermassive black holes using broad iron emission lines (Reynolds, 2003), the extremely hot accretion disk makes it unlikely that these iron lines exist around M 87 or Sgr A*. The EHT, however, provides the resolution to directly observe accretion and variability around M 87 and Sgr A*.

1.3 The Event Horizon Telescope

The EHT is a global network of mm-radio dishes. The 2017 EHT array consisted of 8 telescopes which are listed in Table 1.1, and shown visually in Figure 1.1. By utilizing the techniques of very-long-baseline interferometry (VLBI), the EHT combines the individual telescopes into one radio dish, achieving the highest angular resolution ($20\ \mu\text{as}$) telescope in the world.

In a series of papers Event Horizon Telescope Collaboration et al. (2019a,b,c,d,e,f) (hereafter EHTC I; EHTC II; EHTC VII; EHTC IV; EHTC V; EHTC VI respectively),

Telescope	Location	Median SEFD
ALMA (AA)	Chili	74
APEX (AP)	Chili	4700
JCMT (JC)	Hawaii	10 500
SMA (SM)	Hawaii	6200
LMT (LM)	Mexico	4500
SMT (AZ)	Arizona	17 1000
SPT (SP)	South Pole	19 300

Table 1.1: Telescopes that participated in the 2017 EHT observations, including the median SEFD. These numbers are taken from [EHTC II](#).

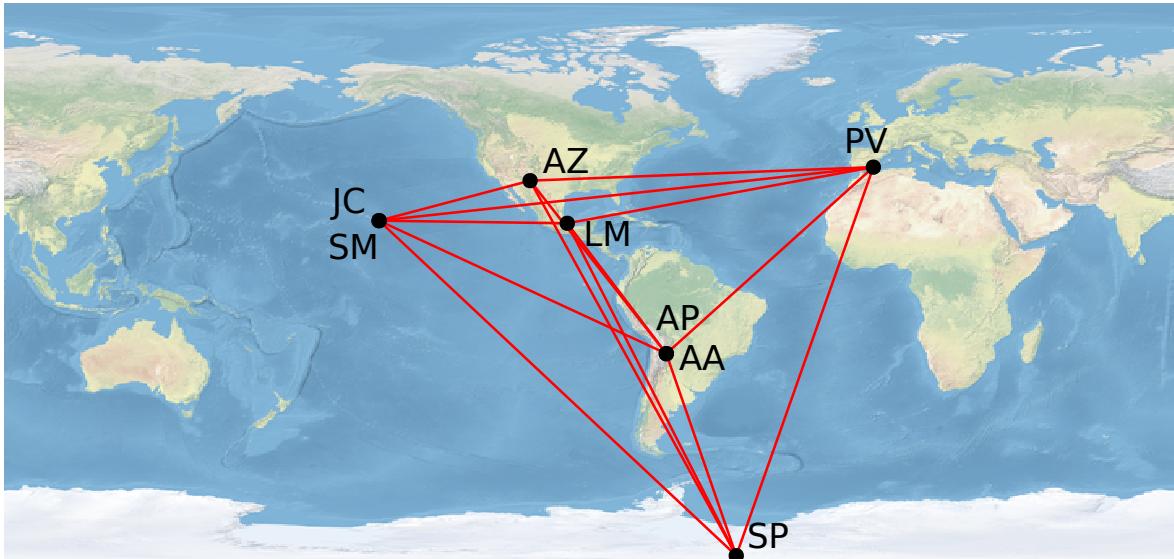


Figure 1.1: World map with the 2017 EHT array stations highlighted.

the EHT published the first image of the black hole at the center of M 87. Follow-up work published in 2021 ([Event Horizon Telescope Collaboration et al., 2021a,b](#), hereafter [EHTC VII; EHTC VIII](#) respectively) produced the first polarized image of a black hole. These results provided the first horizon resolved image of a black hole and direct visual proof that black holes power AGN. Furthermore, the brightening on the bottom is consistent with Doppler beaming due to the relativistic motion of plasma. In the follow-up polarization analysis, the magnetic field structure near the horizon was measured to be largely poloidal

(EHTC VII). By using the inferred polarization structure of the image reconstructions, general-relativistic-magneto hydrodynamical (GRMHD) simulations with large magnetic fluxes, i.e., magnetically arrested disks, were found to be preferred (EHTC VIII).

A primary goal of the first set of EHT papers was to measure the mass of the central black hole. While the EHT analysis in EHTC IV; EHTC VI, produced a ring diameter, its relation to the black hole mass is complicated. Theoretically, the diameter of the ring and the black hole mass should be related. For instance, if the image was dominated by the photon rings described above, the relationship between mass and ring diameter, d , would be $d \approx 2\sqrt{27}GM/c^2$ for a Schwarzschild black hole. However, the direct emission from the accretion disk dominates most of the observed emission (EHTC V) and shifts the observed ring radius. Furthermore, the appearance of the accretion disk is very sensitive to the properties of the emitting electrons (EHTC V). To overcome these theoretical challenges EHTC VI, used an empirical calibration procedure to account for the accretion disk uncertainty.

The calibration procedure started by postulating that the measured ring diameter and black hole mass were related by

$$d = \alpha \frac{GM}{Dc^2}, \quad (1.4)$$

where α is an unknown variable which needs to be estimated, M is the black hole mass, and D is the observer to black hole distance. To estimate α , 100 GRMHD simulations were selected from the EHTC V library. From these simulations, simulated data matching the properties of the 2017 M 87 observations was constructed. For each simulated data set, the measured ring diameter was computed using two techniques. The first used the ensemble of image reconstructions from the three imaging pipelines used in EHTC IV. Given these images, the feature extraction tool REx was used to measure the ring diameter. The second technique was to directly model the observed visibilities or closure products using simple geometric crescent models. For the imaging approach, $\alpha = 10.5 - 11$ depending on the imaging pipeline. For the geometric modeling $\alpha = 11.5 - 12.0$. Using this conversion factor, the measured diameter from both the imaging and geometric modeling implied that M 87 had a mass of $6.5 \pm 0.7 \times 10^9 M_\odot$.

This result constituted the first horizon resolved mass measurement of a black hole outside of our galaxy. Furthermore, this mass measurement settled a long-standing debate on the mass of M 87. The conflict was based on two independent methods using stellar (Gebhardt et al., 2011) or gas (Walsh et al., 2013) motion. The EHT result favored the stellar mass, and implied that certain assumptions about the dynamical state of gas around low-z AGN need to be revisited (Jeter et al., 2019). Moreover, reconciling the stellar and

gas mass measurements suggests that M 87’s outer accretion disk is misaligned ([Jeter & Broderick, 2021](#)).

1.4 Themis: An EHT Bayesian Parameter Estimation Framework

If the EHT is going to detect deviations from GR, understanding uncertainty is paramount. While the EHT can achieve extraordinary resolution, its coverage and dynamic range are relatively poor ([EHTC III](#); [EHTC IV](#)). As a result, any imaging or modeling of the data will be highly uncertain.

Radio astronomy has a long history of overcoming these imaging uncertainties through a variety of techniques. The most prominent of said techniques are deconvolution algorithms such as the venerable CLEAN procedure ([Hogbom, 1974](#); [Schwarz, 1978](#); [Clark, 1980](#); [Schwab, 1984](#)). Additionally, due to increases in computational power, forward modeling approaches such as *regularized-maximum-likelihood* (RML) methods are now becoming popular ([Frieden, 1972](#); [Gull & Daniell, 1978](#); [Narayan & Nityananda, 1986](#); [Chael et al., 2016a, 2018a](#); [Akiyama et al., 2017a](#)). RML methods work by forward modeling the image and then comparing it to the observed visibilities. These methods are very flexible and can be automated more easily than CLEAN. Furthermore, their development was critical for the first images of M 87. However, both RML and CLEAN rely on tuning hyperparameters, e.g., regularizers. Currently, there is no applicable mathematical framework to assess which regularizers work well with the present EHT array¹. Instead, a series of parameter surveys and heuristics are employed ([EHTC IV](#)). Furthermore, both RML and clean methods only produce a single image per set of regularizers, while there are infinitely many images consistent with the data. That is to say, RML and CLEAN methods cannot assign uncertainty to images. Quantifying uncertainty motivates the use of Bayesian techniques for imaging.

1.4.1 Bayesian Inference

Unlike traditional CLEAN or RML based imaging, Bayesian models do not produce a single image, but rather a distribution. With this distribution, image uncertainty can be quantified. For instance, error bars can be placed on quantities of interest, such as the robustness of subdominant objects, or image features such as the ring diameter.

¹Although see [Akiyama et al. \(2017a\)](#) for a potential solution

Bayesian methods assume that any model, M , and its parameters, θ , are inherently uncertain. Rather than describe model parameters as a single number to be “discovered”, they are thought of as probabilistic statements on certainty and described by a distribution. The probability distribution of the model parameters before a measurement is called the prior $p(\theta|M)$. To connect the model to the data, we require the likelihood $\mathcal{L}(V|\theta, M)$. The likelihood denotes the probability of observing the data, V , given the model and parameters. In addition, the likelihood acts as a penalty function, moving us from the prior distribution to the posterior through Bayes theorem

$$\pi(\theta|V, M) = \frac{\mathcal{L}(V|\theta, M)p(\theta|M)}{p(V|M)}. \quad (1.5)$$

The term in the denominator is typically known as the marginal likelihood or evidence, and is needed to ensure that the posterior is normalized. The posterior describes the level of certainty of the model parameters after the observations. If the observations are informative, the posteriors will appear different from the prior. However, if our observations are not informative, the resulting distribution will appear similar to the prior. In the large data limit, when the number of observations dominates over the number of parameters, the Bernstein-von Mises theorem states that the Bayesian estimates converge to the frequentist maximum likelihood estimates. However, this theorem does not hold in several circumstances. For instance, if the model is misspecified², then the Bayesian credible intervals cannot be interpreted as frequentist confidence intervals ([Kleijn & van der Vaart, 2012](#); [Grünwald & van Ommen, 2018](#)).

The RML imaging problem can, in some sense, be interpreted as a restricted version of Bayesian inference. In this setting, the regularizers are replaced by priors³. The image estimate is then given by the maximum a posteriori (MAP). However, the MAP does not give any information about the uncertainty of the reconstruction. Fully Bayesian methods are interested in the distribution of consistent images. Generally, finding this distribution is difficult and computationally expensive. The next section describes the approach taken in this thesis, Markov Chain Monte Carlo.

1.4.2 Markov Chain Monte Carlo

The main goal of Bayesian inference is to find the posterior. Using the posterior we can then calculate quantities of interest through expectations i.e., integrals of the posterior

²“All models are wrong, but some are useful” ([Box, 1976](#))

³However, these regularizer functions may not be probability distributions, meaning that the posterior may not exist.

with some function f :

$$\mathbb{E}_\pi [f] = \int f(\theta) \pi(\theta|V, M) d\theta. \quad (1.6)$$

Computing these expectations are computationally challenging for two reasons. First, θ is often a high-dimensional vector, making the integrals very expensive to compute using standard Riemannian methods. Second, expectations require computing the evidence $p(V|M)$, which is difficult. Monte Carlo methods were devised to solve this specific problem. Monte Carlo methods work by sampling from the probability distribution you wish to explore. Using these N samples, θ_i , one can form the empirical distribution, or Monte Carlo approximation of $\pi(\theta|V, M)$:

$$\pi(\theta|V, M) \approx \frac{1}{N} \sum_{i=1}^N \delta(\theta - \theta_i). \quad (1.7)$$

The integrals in (1.6) are then replaced by simple summations:

$$\mathbb{E}_\pi [f] = \frac{1}{N} \sum_{i=1}^N f(x_i). \quad (1.8)$$

One can then show that this approximation converges to the true value at a rate of $1/\sqrt{N}$, regardless of dimension (Robert & Casella, 2013).

However, directly sampling from the posterior is typically intractable. Several techniques exist to solve this problem, such as rejection sampling, importance sampling, nested sampling, and Markov Chain Monte Carlo (MCMC). Rejection and importance sampling are known to scale poorly in dimension due to the difficulty of choosing an approximate bounding/importance distribution (Robert & Casella, 2013). Nested sampling (Skilling, 2006), is a modern technique that attempts to solve the problem by sampling from increasing likelihood bounding regions. Nested sampling works for multimodal distributions with complex correlated structures (Feroz et al., 2009). However, current nested sampling algorithms struggle in high dimensions, scaling as $\mathcal{O}(d^3)$ (Handley et al., 2015).

MCMC methods (Metropolis et al., 1953; Hastings, 1970) are simple to implement and are widely used in statistics and astronomy. Furthermore, they provide an asymptotically exact method to produce samples from the posterior at the cost of them being correlated. The theoretical properties of MCMC methods have been researched for decades, providing several tools to assess their performance and quantify their errors (see Brooks et al., 2011, for a review). As well, certain MCMC methods should theoretically scale well with dimension (Beskos et al., 2013, argues the scaling for Hamiltonian Monte Carlo is $\mathcal{O}(d^{4/3})$).

Almost all current MCMC methods use a form of the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970). The MH algorithm depends on two ingredients: the complex distribution you wish the sample from π and a proposal distribution q . Given an initial position, θ , the MH algorithm uses the proposal distribution $q(\theta'|\theta)$ to suggest a new position θ' . This new position is accepted with probability

$$\alpha = \min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right). \quad (1.9)$$

If the proposal distribution is identical to π , then $\alpha = 1$, reducing to perfect sampling. Almost all current MCMC methods, e.g., Hamiltonian Monte Carlo, Gibbs Sampling, Metropolis-Adjusted Langevin, Affine Invariant MCMC, are examples of MH with a specific proposal distribution. This proposal distribution is the main object to consider when implementing an MCMC method. If the proposal distribution is inefficient, MCMC methods may not converge in feasible timescales.

1.4.3 Themis

Bayesian inference requires many different ingredients. For the EHT, this was a technological hurdle. To solve this challenge, Broderick et al. (2020) developed a Bayesian parameter estimation framework called THEMIS, specifically targeted for the EHT. THEMIS is a flexible framework written in C++ and MPI to enable massive parallelization. A typical THEMIS program is built from “blocks”, that encapsulate different aspects of Bayesian programming. These blocks are written in terms of the following C++ classes:

1. **Data:** Data class, that describes the different data products of the EHT.
2. **Model:** Model class that parameterizes some image that produces visibilities, e.g. Gaussian’s, crescents, semi-analytical models.
3. **Priors:** Prior class
4. **Likelihood:** Connects the model and data, and describes the observational error.
5. **Sampler:** MCMC/Sampling framework for efficient posterior exploration

Each of these components is described in detail in Broderick et al. (2020). The benefit of this compartmentalized approach is that THEMIS is easily extendable. For example, including new models only requires that the user understands the model class. As a

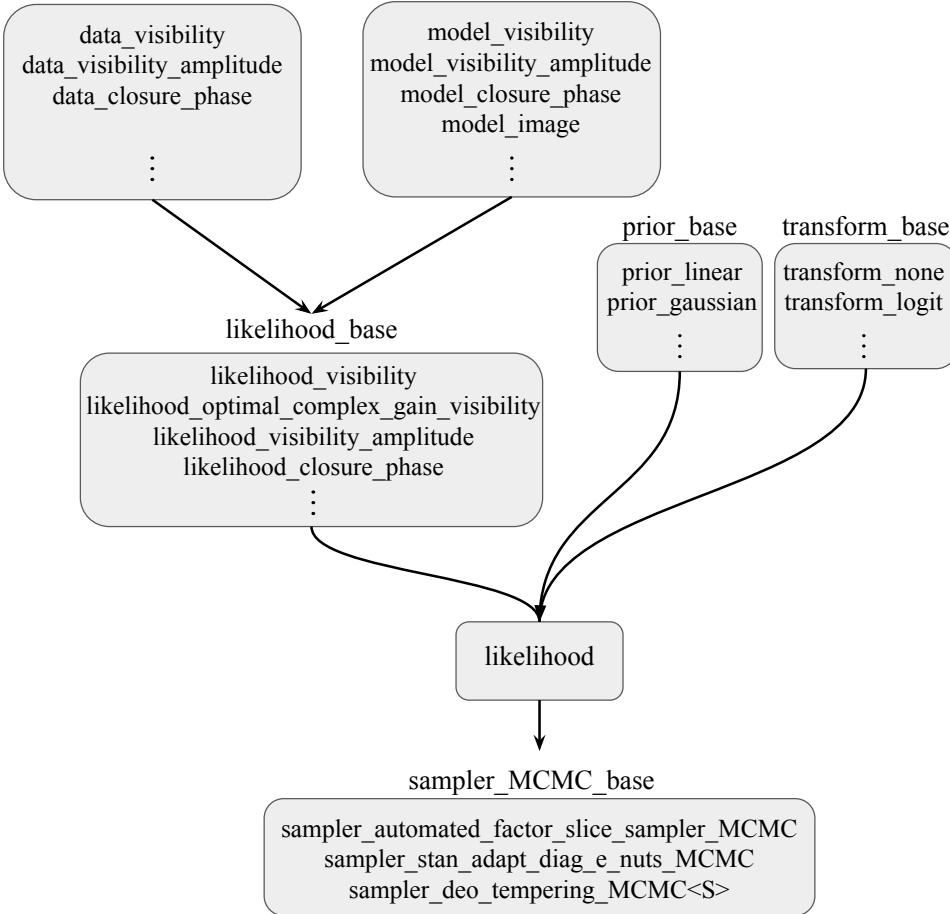


Figure 1.2: The layout of THEMIS. Each box defines an interface to the data, model, likelihood, prior, and transform classes. As well, each interface has a number of child classes, with only a small portion being shown in the figure. These compartments are eventually merged into a large likelihood class that computes the un-normalized posterior. This likelihood can then be passed to a sampler that will produce draws from the posterior using MCMC methods. This compartmentalization allows users to easily extend the parts they are interested in.

result of this flexibility, THEMIS was the driving force in the first EHT results on M 87 (EHTC VI; EHTC VII), 3c279 (Kim et al., 2020), and the historical variability analysis of M 87 (Wielgus et al., 2020). Furthermore, THEMIS has been used to produce the first fully

Bayesian radio-image reconstructions (Broderick et al., 2020). In this thesis, the results from Chapters 4, 5, and 6 use the THEMIS framework.

1.5 Summary

This thesis explores the capabilities of the current EHT to constrain the structure of spacetime on horizon scales. This will be accomplished by analyzing the shape of the black hole image, and variability of Sgr A* during flares. General relativity predicts that the shadow should be symmetric given the inclination angle of the black hole M 87 (Mertens et al., 2016). A robust detection of ellipticity in the image reconstructions of M 87 could signal new gravitational physics on horizon scales.

The first part of the thesis develops a method called *variational image domain analysis* (VIDA) (Chapter 2) that extracts features such as ring ellipticity from image reconstructions. Feature extraction is required since imaging is non-parametric. To demonstrate VIDA’s flexibility, we analyzed thousands of reconstructions from previous EHT simulated datasets to recover image features such as diameter, orientation, and ellipticity. Chapters 3–5 use VIDA to constrain the ellipticity of M 87 using RML and Bayesian imaging methods.

In Chapter 3, we analyze whether the RML imaging pipeline used in EHTC IV is sensitive to ring ellipticity. We find that the set of imaging hyperparameters used for M 87 is unable to recover ellipticity. Even for simple geometric models, the true ellipticity is not recovered. To calibrate for the imaging bias, we then image 550 GRMHD images, using the M 87 hyperparameters, and find that GRMHD simulations with an 3 : 2 axis ratio are consistent the M 87 eht-imaging results. Measuring the sensitivity to additional ellipticity that could occur from non-GR effects, we find that simulations with intrinsic axis ratios of 2 : 1 are consistent with the M 87 eht-imaging results.

The results of Chapter 3 are a symptom of standard imaging techniques inability to assign a statistically meaningful uncertainty to image reconstructions. However, with THEMIS, Bayesian imaging of radio astronomy datasets is now possible. To enable Bayesian imaging requires exploring high dimensional multi-modal distributions. Chapter 4 describes the next generation samplers implemented in THEMIS. These samplers include a state-of-the-art Hamiltonian Monte Carlo (HMC) sampler using the no u-turns (NUTS) adaptation scheme. While NUTS works very well for single-mode distributions, it struggles to move between well-separated modes. By combining HMC with a non-reversible parallel tempering scheme, we allow for a global exploration of parameter space. These improvements have enabled polarized Bayesian imaging, which requires efficient sampling of models with 300 – 500 parameters.

In Broderick et al. (2020), a novel Bayesian imaging technique was applied to simulated data based on the 2017 EHT coverage of M 87. Chapter 5 extends this analysis to the ellipticity measurement of M 87. By combining the image posteriors from THEMIS with feature extraction techniques from VIDA, we produce the first-ever Bayesian ellipticity estimates of M 87. This is done in two steps. First, we validate the Bayesian ellipticity method on the simulated ellipticity tests from Chapter 3, that the RML methods failed to recover. Second, we analyze the M 87 data and find an axis ratio $\lesssim 5 : 4$ across all days. Furthermore, we show that this ellipticity measurement is consistent with the expected ellipticity from GRMHD simulations.

Future observations of dynamical phenomena provide a powerful way to probe space-time properties. Sgr A*, flares 1–3 times a day depending on the wavelength. These flares could arise through magnetic reconnection events in the accretion flow, forming a hot spot. In Chapter 6, we construct a semi-analytical model that includes the effects of shearing as a spot moves along the accretion flow. We then explore the ability of the 2017 EHT to recover said hot spots. Even including significant systematic uncertainties, such as thermal noise, diffractive scattering, and background emission due to an accretion disk, we were able to recover the hot spot and black hole spin to sub-percent precision. Moreover, by observing multiple flaring events, we show how the EHT could be used to *tomographically map spacetime*. This provides new avenues for testing relativistic fluid dynamics and general relativity near the event horizon of supermassive black holes.

Chapter 2

Variational Image Domain Analysis for the Event Horizon Telescope

2.1 Introduction

Generating quantitative measurements about intrinsic radio images from very long baseline interferometry (VLBI) is a computationally and theoretically difficult task. In practice, the small number of stations participating in the Event Horizon Telescope (EHT), and thus sparse coverage in the u - v plane, results in variety of potential image structures. As a result, the process of reaching quantitative conclusions about image features requires significant additional analysis. For the EHT analysis of the 2017 M 87 data, this has taken two, complementary forms ([EHTC I](#)).

The first is a traditional Bayesian parametric modeling approach ([EHTC VI](#)). Here, simple geometric models are fit to the visibility data, and direct quantitative inferences about the image properties are encoded within the model, e.g., ring diameter. However, to give reliable estimates, these geometric models need to provide good approximations to the true on-sky image. Therefore, while geometric models give direct measurement of relevant features, they require a priori knowledge of the image structure.

The second approach is non-parametric, and usually referred to as “imaging” ([EHTC IV](#)). This category of methods is broadly defined, and includes deconvolution algorithms like CLEAN ([Hogbom, 1974](#); [Schwarz, 1978](#); [Clark, 1980](#); [Schwab, 1984](#)) and forward modeling approaches like “maximum entropy” ([Frieden, 1972](#); [Gull & Daniell, 1978](#); [Narayan & Nityananda, 1986](#)), regularized maximum likelihood (RML) analyses ([Chael et al., 2016a](#),

2018a; Akiyama et al., 2017a,b), and Bayesian imaging (Broderick et al., 2020). For the EHT, the output of imaging is an ensemble of image reconstructions that reproduce the observed visibility data Event Horizon Telescope Collaboration et al. 2019d. These imaging methods have the significant advantage that they are extremely flexible, and therefore are reasonably expected to cover the “truth”. However, unlike parametric modeling, imaging methods do not give direct quantitative measurements of the image features of interest, e.g., ring diameter, width, orientation, etc. Therefore, reaching quantitative conclusions about the image properties requires an additional processing step, which we call “feature extraction”. It is this final step that is the subject of this chapter.

Feature extraction is similar to the geometric/parametric modeling above, but is applied in the image domain rather than visibility domain. However, it differs in one important respect: by virtue of being performed *after* imaging, the class of applicable “models”, i.e., image features to be measured, is already well known. For example, in EHTC IV and EHTC VI, quantitative image features were extracted by the algorithm “ring extractor” REx (Chael, 2019). However, REx is only applicable to images that have a dominant ring-like feature. In general, images from the EHT can have a complex structure and are dependent on the intrinsic source. For instance, the active galactic nucleus 3c 279 displays a jet morphology which while poorly described by a ring, can be described by a set of Gaussians (Kim et al., 2020).

One possible approach to feature extraction is to “template” relevant image features using a transformation. For example, the Hough transform (Hough, 1964; Duda & Hart, 1972), is used to extract rings and other shapes from images using template matching. A related method is to approximate the complicated image reconstruction with parametric templates that describe the features of interest. This idea is more akin to the parametric/-geometric modeling approach of visibility data in EHTC VI and is the approach taken in this chapter.

Any comparison requires a suitable quality metric, i.e., objective function. Because the total flux is typically an arbitrary rescaling, and the image brightness is positive definite, there is a natural identification between the flux-normalized image and probability distribution. This motivates the use of “divergences” as an extremely flexible class of objective functions for comparing images; for this reason divergences have been used extensively in image processing (e.g., Goudail et al., 2004; Choi & Lee, 2003; Aherne et al., 1998).

We adopt a method similar to variational inference (Blei et al., 2017), in which complicated distributions are approximated by simple parametric forms. The optimal parameters are estimated via the minimization of an appropriate divergence. In this chapter, we develop a number of appropriate parametric forms and explore the performance of a variety

of divergences for application to image feature extraction. Therefore, we call this method *variational image domain analysis*, or VIDA. The VIDA algorithm has been implemented in the open source package `VIDA.jl`¹ written in Julia (Bezanson et al., 2017).

There are two main reasons that imaging reconstruction followed by feature extraction is advantageous when compared to directly fitting simple geometric models to the data. First, choosing the correct geometric model is difficult before imaging, meaning imaging is the first step in both methods. Second, fitting simple geometric models can “underfit” the data leading to biased results. For instance, in EHTC VI additional “nuisance” Gaussians were required to obtain a reasonable reduced chi-square. On the other hand, Bayesian imaging techniques (e.g., Broderick et al. (2020) and Chapter 5) make minimal assumptions about source structure. Combining the Bayesian imaging posterior with VIDA then provides a deterministic map from image to feature posteriors. VIDA’s feature posteriors can then be compared to the geometric modeling results, testing whether the features are robust across methods. Therefore, VIDA fills a gap in the EHT modeling pipeline and is generic, unlike current EHT tools.

The layout of this chapter is as follows: In Section 2.2, we present the details of VIDA. Namely we detail, the different types of templates implemented, and the objective function used to find the best approximation to the true image. Section 2.3 applies VIDA to a variety of ring-like image reconstructions from the test set of EHTC IV and compares the results to REx. This is an empirical demonstration that we can recover the optimal template, even through the objective function is non-convex. In Section 2.4, we demonstrate VIDA’s flexibility by applying it to non-ring images from the test set of EHTC IV. Finally, the conclusions are detailed in Section 2.5.

2.2 Variational Image Domain Analysis

The critical insight behind VIDA is that images (sans polarization and modulo total flux) and probability densities are in one-to-one correspondence. Namely, images are point-wise positive and integrable. Probability divergences are a natural class of objective functions used to compare two probability distributions. Furthermore, they have been used in image feature extraction and similarity measures before.

VIDA consists of three ingredients:

1. **Image $I(\alpha, \beta)$** whose features we want to extract

¹<https://github.com/ptiede/VIDA.jl>

2. **Template** (or approximate image) that parameterizes the features of interest, e.g. ring radius
3. A **divergence**, i.e. the objective function we minimize

Each of these building block have independent abstract types in `VIDA.jl` enabling users to easily add additional image templates, and divergences². In [Section 2.2.1](#) and [2.2.2](#) we will review the templates and divergences currently implemented in `VIDA` respectively.

2.2.1 Image Templates in `VIDA.jl`

The choice of template used will depend on the structure of the image. For example, the images of M 87 from [EHTC IV](#) are ring-like, while the reconstructions of 3c 279 from [Kim et al. \(2020\)](#) can be described by several Gaussian brightness distributions. In this section, we present the various templates that are implemented in `VIDA.jl`.

Gaussian template

To model a source of compact flux we include an asymmetric Gaussian template. The parameters of the Gaussian template are:

1. The size, $\sigma = \sqrt{\sigma_a \sigma_b}$, where $\sigma_{a,b}^2$ are the variances in the principal directions of the Gaussian.
2. The ellipticity, $\tau = 1 - \sigma_b/\sigma_a$, measures the ellipticity of the Gaussian and we assume $\sigma_a > \sigma_b$.
3. ξ , rotation angle (relative to the Gaussian center) of the principal axes measured east of north.
4. x_0, y_0 , the center of the Gaussian.

²See the documentation at <https://ptiede.github.io/VIDA.jl/dev/> for a tutorial on how to add additional templates

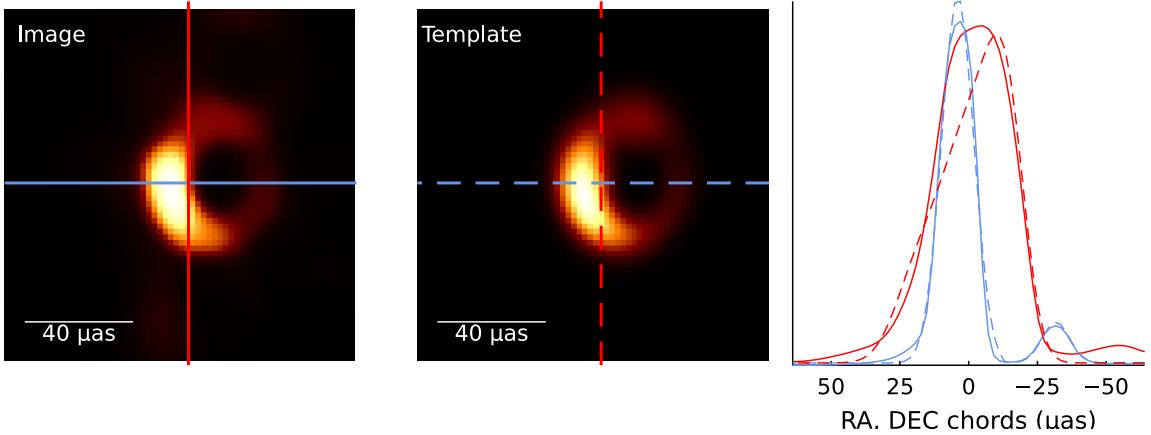


Figure 2.1: An example of VIDA run. *Left:* image reconstruction of a GRMHD simulation from [EHTC V](#); [EHTC IV](#). *Middle:* VIDA reconstruction using the `CosineRing{1,4}` template and Bh divergence. *Right:* vertical (red) and horizontal (blue) chords through the center of light of the truth image. The dashed lines are for the VIDA optimal template, and solid for the reconstruction. This plot can be made in the `VIDA.jl` package using the `triptic` function and is used to assess the quality of the template approximation.

Disk template

In [Section 2.3](#) we will test `VIDA.jl` on a number of synthetic data tests. One of the test images is a disk. To approximate disks we use the template:

$$f_{\text{Disk}}(r; r_0, \alpha, x_0, y_0) = N \begin{cases} 1 & r < r_0 \\ \exp(-(r - r_0)^2 / 2\alpha^2) & r > r_0, \end{cases} \quad (2.1)$$

where r_0 is the radius of the flat disk, α controls the smoothness of fall off and N ensures the template is normalized. The radial distance, r , is relative to the center x_0, y_0 and N is the normalization. When $r_0 = 0$ this template reduces to a symmetric Gaussian with standard deviation α .

Ring templates

One of the principal quantities of interest in images of M 87 is the ring diameter, d_0 , since it is related to the mass of the central black hole. Additionally, the ring is expected to have some thickness, w , due to the emitting material around the black hole. The

simplest template would be a circular Gaussian ring with some thickness. Doppler boosting, however, can cause the emission to appear asymmetric. To model this a slash can be added to the ring template. Additionally, the ring itself does not have to be circular. Ring ellipticity could occur from e.g., the emitting material not being azimuthally symmetric around the black hole. As well, due to the sparse coverage of the EHT array and imaging algorithms, ellipticity may be introduced into the reconstructions. Consolidating each of these features into a template parameter, we get the following:

- d_0 : the geometric mean of the semi-major, a , and semi-minor, b , axis $d_0 = 2\sqrt{ab}$ which is related to the area of the ellipse, $\pi(d_0/2)^2$.
- τ : the ellipticity of the ellipse, $\tau = 1 - b/a$
- ξ_τ : the position angle of the semi-major axis measured east of north.
- w : the width of the Gaussian ring, defined to be the full width half max (FWHM) of the Gaussian, i.e. $w = 2\sqrt{2 \log 2}\sigma$, where σ is the standard deviation.
- s : the strength of the slash described in equation (2.4).
- ξ_s : the position angle of the slash measured east of north.
- (x_0, y_0) : the center of the ring.

The functional form of the template is given by:

$$h_\theta(x, y) = S(x, y; s, \xi) \exp \left[-\frac{(d_\theta(x, y))^2}{2\sigma^2} \right], \quad (2.2)$$

where $S(x, y; s, \xi)$ is the slash function and $d_\theta(x, y)$ is the minimum distance between the ellipse with parameters $\theta = (d_0, \tau, x_0, y_0)$ and the point x, y . If $\tau = 0$, $d_{\tau=0}(x, y) = |\|\mathbf{x} - \mathbf{x}_0\| - d_0/2|$. However, for an ellipse there is no analytical equation. Instead one has to numerically minimize the function,

$$L(x, y, e_x, e_y) = \|(x - e_x, y - e_y)\|, \quad (2.3)$$

subject to the constraint that e_x, e_y are points on the ellipse with parameters d_0, τ, ξ_τ .

For the slash function S , we use a first order cosine expansion in azimuthal angle ϕ around the center x_0, y_0 :

$$S(x, y; s, \xi_s) = N_0 [1 + s \cos(\phi - \xi_s)], \quad (2.4)$$

where N_0 is a normalization factor to ensure the template is unit normalized. To prevent image flux from becoming negative, we restrict $s \in [0, 1]$. In the VIDA.jl package, this template is called `GeneralGaussianRing` (GGR) and an example reconstruction using said template is shown in [Figure 2.1](#).

Additionally, VIDA.jl has a number of other ring-like templates currently implemented:

- `GaussianRing`: Symmetric Gaussian ring with constant azimuthal intensity (i.e. GGR with $\tau, s = 0$)
- `SlashedGaussianRing`: Symmetric Gaussian ring with azimuthal slash described by [Equation 2.4](#) (i.e. GGR with $\tau = 0$)
- `EllipticalGaussianRing`: Elliptical Gaussian ring with constant azimuthal flux (i.e. GGR with $s = 0$)
- `TIDAGaussianRing`: GGR template where the slash and ellipticity position angle are either aligned or anti-aligned.

We also include a more general version of the GGR called the `CosineRing{N,M}`. This template is similar to the GGR template but where the width, σ , and slash function (2.4) are replaced by a higher order cosine expansion in azimuthal angle ϕ :

$$S_M(\phi; \mathbf{s}, \boldsymbol{\xi}^{(s)}) = 1 - \sum_{m=1}^M s_m \cos[m(\phi - \xi_m^{(s)})], \quad (2.5)$$

$$\sigma_N(\phi; \boldsymbol{\sigma}, \boldsymbol{\xi}^{(\sigma)}) = \sigma_0 + \sum_{n=1}^N \sigma_n \cos[n(\phi - \xi_n^{(\sigma)})], \quad (2.6)$$

where \mathbf{s} , $\boldsymbol{\sigma}$, $\boldsymbol{\xi}^{(\cdot)}$, are vectors with the cosine expansion coefficients of the slash, standard deviation, and angular offset. We can reproduce the `GeneralGaussianRing` template by setting $M = 1$ and $N = 0$ in (2.5) and (2.6) respectively. This template can be used if image has a ring-like feature that has a bumpy azimuthal profile.

Background Templates

We found that many image reconstructions had a diffuse amount of flux throughout the image due to poor dynamic range from sparse coverage and regularization effects. To

model the background, we added a constant intensity template. This template is typically required to be included in any analysis to ensure reliable feature extraction.

A more complex background structure is also included with the `DiffuseBack` template. This decomposes the background into a uniform grid of intensities that are blurred with some Gaussian kernel. Typically we find a very coarse 2×2 or 3×3 grid works well.

Composite Templates

A general image reconstruction from a VLBI observation may have multiple image features. As such, `VIDA.j1` allows the user to combine multiple features into composite templates, where each individual component is given a relative flux³. An example of this is shown in [Section 2.4.1](#) where three Gaussian templates are used to model the reconstructed images.

2.2.2 Probability Divergences

As mentioned above, `VIDA.j1` uses an analogy between images and 2-D probability distributions to motivate the use of divergences as objective functions. Divergences form a general measure of similarity between two distributions. A divergence can be thought of as a functional $F_q[p] = \mathcal{D}(p||q)$, comparing a the distribution p (template) to a reference q (image), and is required to be non-negative, $\mathcal{D}(p||q) \geq 0$, and non-degenerate, $\mathcal{D}(p||q) = 0$ if and only if $p = q$. Note that this definition is more general than a metric. Namely, a divergence does not have to be symmetric, i.e. $\mathcal{D}(p||q) \neq \mathcal{D}(q||p)$ or satisfy the triangle equality.

One of the most well-known divergences is the Kullback-Leiber (KL) divergence ([Kullback & Leibler, 1951](#)) or relative entropy,

$$\text{KL}(p||q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) d^2x. \quad (2.7)$$

One issue with the KL divergence is its definition when the support of q and p differ, i.e. if $q(x) = 0$. In this case we set the contribution to the integral to be zero.

In addition to the KL divergence `VIDA.j1` also includes the Bhattacharyya divergence (Bh) ([Bhattacharyya, 1943](#)),

$$\text{Bh}(p||q) = -\log \int \sqrt{p(x)q(x)} d^2x. \quad (2.8)$$

³Note the absolute or total flux of the image is not recoverable since we renormalize each image to have unit flux.

The Bh divergence is related to a well known metric on probability spaces, the *Square Hellinger Distance* :

$$H(p, q) = \frac{1}{2} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 d^2x = 1 - \text{Bh}(p||q). \quad (2.9)$$

Therefore, the minimizing the Bh divergence is simply a least squares fit in the space of the square root of distributions.

In this chapter we will present the results from optimizing the Bh divergence for two reasons. First, we found that while the KL and Bh divergence produce near-identical results, the Bh divergence required $\sim 25\%$ less evaluations to converge. Second, the Bh divergence has preferable theoretical properties compared to the KL divergence. Namely, it is well defined when the image pixels have zero flux, and is symmetric.

2.2.3 Optimizing the Divergence

A problem when using probability divergences is that they give non-convex, non-linear optimization problems. Furthermore, the nature of the problem will change if the template changes, making an analytic analysis difficult. Therefore, to extract the globally optimal template, we turned to heuristic global optimizers, such as Genetic/Evolutionary strategies (see [Das & Suganthan, 2011](#), for a review) and simulated annealing ([Goffe, 01 Oct. 1996](#)). For this chapter we used the Julia package BlackBoxOptim.jl⁴ ⁵. BlackBoxOptim.jl uses natural-evolution and differential evolution strategies to perform a stochastic search of the parameter space. In the next section, we will validate that our chosen optimizer is able to reliably recover the optimal template.

2.2.4 REx and relating to VIDA parameters

The first step in REx (see [EHTC IV](#), for details) is to identify the dominant ring in the image. Given a center position (x, y) , REx samples the image in radius r and azimuthal angle θ , obtaining a intensity map $I(r, \theta|x, y)$. Then for that central map, the radius of the ring is taken as the azimuthally averaged peak brightness:

$$\begin{aligned} r_{\text{pk}}(\theta|x, y) &= \text{argmax}[I(r, \theta|x, y)]_r \\ \bar{r}_{\text{pk}} &= \langle r_{\text{pk}}(\theta|x, y) \rangle_{\theta \in [0, 2\pi]}. \end{aligned} \quad (2.10)$$

⁴<https://github.com/robertfeldt/BlackBoxOptim.jl>

⁵VIDA.jl also has interfaces to other Julia optimization packages, such as Optim.jl and CMAESEvolutionStrategy.jl

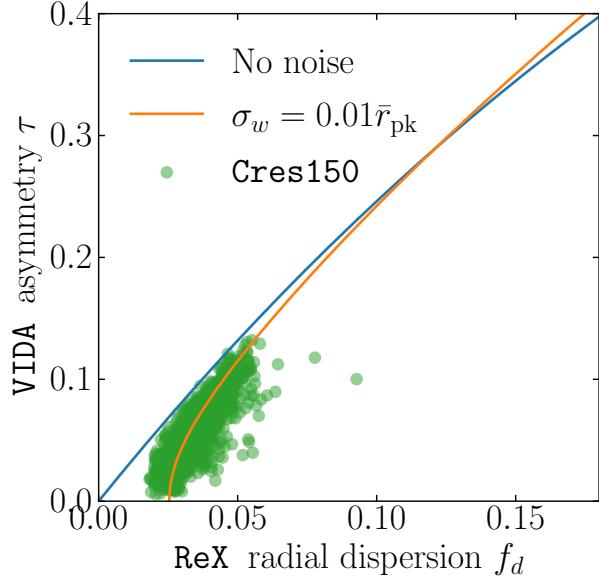


Figure 2.2: Comparison of the REx’s fractional radial dispersion and VIDA’s τ ellipticity parameter. The blue curve shows the conversion for the case of an ellipse, the orange curve is an ellipse with a constant white noise fluctuation in the found radius, with a variance of $0.01^2 \bar{r}_{\text{pk}}^2$. The green points are the results of fitting the Cres150 top set in Section 2.3 with both REx and VIDA.

This provides a different “radius” for every point (x, y) in the image. To identify the ring center, the fractional radius spread is minimized:

$$(x_0, y_0) = \operatorname{argmin} \left[\frac{\sigma_{\bar{r}}(x, y)}{\bar{r}_{\text{pk}}(x, y)} \right], \quad (2.11)$$

where $\sigma_{\bar{r}}(x, y) = \langle (r_{\text{pk}}(\theta|x, y) - \bar{r}_{\text{pk}})^2 \rangle$, is the radial dispersion. The diameter of the ring is then:

$$d = 2\bar{r}_{\text{pk}}(x_0, y_0). \quad (2.12)$$

To relate this to VIDA’s definition we consider an ellipse with semi-major axis a and semi-minor axis b . VIDA parameterizes this ellipse with $d_0 = 2r_0 = 2\sqrt{ab}$ and $\tau = 1 - b/a$. The relationship between r_0 and r_{pk} is given by

$$\bar{r}_{\text{pk}} = \frac{r_0}{\sqrt{1-\tau}} \frac{1}{2\pi} \int_0^{2\pi} \sqrt{1 - e^2(\tau) \sin^2(\theta)} d\theta = \frac{2}{\pi} \frac{r_0}{\sqrt{1-\tau}} E(\epsilon(\tau)), \quad (2.13)$$

where $E(x)$ is the complete Elliptic integral of the second kind and $\epsilon(\tau) = \sqrt{1 - (1 - \tau)^2}$ is the orbital eccentricity.

REx's measure of circularity is provided by the radial fractional dispersion:

$$f_d = \frac{\sigma_{\bar{r}}}{\bar{r}_{\text{pk}}}. \quad (2.14)$$

To compare **REx**'s ellipticity measure, f_d , to **VIDA**'s, it can be shown that

$$f_d(\tau) = \frac{\sqrt{1 - \epsilon(\tau)^2 - 4/\pi^2 E(\epsilon)^2}}{\sqrt{1 - \tau}} = \frac{\sqrt{(1 - \tau)^2 - 4/\pi^2 E(\epsilon)^2}}{\sqrt{1 - \tau}}. \quad (2.15)$$

[Equation 2.14](#) is then used to convert this to a fractional diameter spread. Using linear interpolation we invert the function achieving a map from f_d to τ . One important thing to note is that this conversion assumes that the image is a perfect ellipse. In general, this will not be true for the image reconstructions. Therefore, **REx**'s reported ellipticity may be larger than the **VIDA** measurement. To model the increase we consider r_{pk} modified by a white noise ε_θ term with dispersion proportional to the average radius $\langle \varepsilon_\theta \varepsilon_{\theta'} \rangle = \sigma_w^2 \bar{r}_{\text{pk}}^2 \delta(\theta - \theta')$. With this noise term, the average peak radius is unchanged since ε has mean 0. However, the additional noise does impact the radial dispersion:

$$\sigma_{\bar{r}} \rightarrow \sigma_{\bar{r}} + 2 \langle \varepsilon_\theta r_{\text{pk}}(\theta) \rangle + \langle \varepsilon_\theta^2 \rangle. \quad (2.16)$$

When we have a circular ring then this just becomes $\langle \varepsilon_\theta^2 \rangle = \sigma_\varepsilon^2 \bar{r}_{\text{pk}}^2$ adding an apparent minimal ellipticity to the image after converting to τ . [Figure 2.2](#) shows the conversion when the ring is elliptical and compares it to the results of the Cres150 top set of [Section 2.3](#), with a conversion curve with noise floor (orange) fit by eye.

The width of the ring is defined by finding the FWHM at a fixed θ ray, and then averaging over θ ,

$$w = \langle \text{FWHM}_r[I(r, \theta|x_0, y_0) - I_{\text{floor}}] \rangle_\theta. \quad (2.17)$$

The flux floor is given by $I_{\text{floor}} = \langle I(r = 50\mu\text{as}, \theta) \rangle_\theta$ and is included to avoid biasing the measurement due to the low level flux present in the image. This is similar to including the constant flux template during the **VIDA** extraction.

To characterize the azimuthal profile of the ring (ξ_s and s for **VIDA**), we consider the azimuthal moments of the ring. Namely, the orientation ξ_s is given by:

$$\xi_s = \left\langle \text{Arg} \left[\int_0^{2\pi} I(r, \theta|x_0, y_0) e^{i\theta} d\theta \right] \right\rangle_{r \in [r_{\text{in}}, r_{\text{out}}]}, \quad (2.18)$$

where $r_{\text{in}} = (d - w)/2$ and $r_{\text{out}} = (d + w)/2$. The strength of the slash is given by

$$s = 2 \left\langle \frac{\left| \int_0^{2\pi} I(r, \theta | x_0, y_0) e^{i\theta} d\theta \right|}{\int_0^{2\pi} I(r, \theta | x_0, y_0) d\theta} \right\rangle. \quad (2.19)$$

Note that the factor of 2 is included to match VIDA's definition ([Equation 2.4](#)).

2.3 Validating VIDA

To validate VIDA we need to analyze two related quantities. First, we need to verify that the objective function, i.e. the Bh divergence ([2.8](#)), is robust to the artifacts that occur in image reconstructions. Namely whether the recovered parameter distribution contains the true value. Additionally, given the complex nature of the optimization problem we need to ensure that the chosen optimizer can recover the global maximum. Our validation procedure will consist of:

1. Selecting an applicable ground truth image I_{truth} (see the left column of [Figure 2.3](#))
2. For each truth image create a simulated EHT observation matching the observation characteristic of the EHT M 87 2017 observations.
3. Create an ensemble of image reconstructions of the simulated observations using the same procedure as [EHTC IV](#).
4. Apply VIDA to each image reconstruction and compare the inferred results to the ground-truth parameters, and the reconstruction technique REx used in [EHTC IV](#); [EHTC VI](#).

2.3.1 Step 1: Selecting Ground Truth Images

To validate VIDA, we applied it to a subset of the test set from [Event Horizon Telescope Collaboration et al. \(2019d\)](#). The sources we considered are shown in [Figure 2.3](#), and consist of two geometric crescents and a general-relativistic-magneto-hydrodynamical (GRMHD) simulation from [Event Horizon Telescope Collaboration et al. \(2019e\)](#). The geometric crescent model is described by:

$$I(r, \theta) = I_0 (1 - s \cos(\theta - \xi)) \frac{\delta(r - r_0)}{2\pi r_0}. \quad (2.20)$$

The infinitely thin ring is then convolved with a circular Gaussian with FWHM $10\ \mu\text{as}$. Two orientations (measured east-of-north) $\xi = 180^\circ$ and $\xi = 150^\circ$ are considered in this chapter, and are shown in the left and middle panels of [Figure 2.3](#). These two crescent orientations are denoted by Cres180 and Cres150. For both orientations we took $r_0 = 22\ \mu\text{as}$, $s = 0.46$, and $I_0 = 0.6\ \text{Jy}$. After blurring the ring, it is important to note that the effective radius (the intensity peak) is smaller than the original radius of the ring (see [EHTC IV](#)). The amount the diameter is biased inwards is given approximately by:

$$d_{\text{blur}} = d_{\text{true}} - \frac{1}{8 \ln 2} \frac{\alpha^2}{d_{\text{true}}}, \quad (2.21)$$

where d_{true} is the diameter of the non-convolved ring ($44\ \mu\text{as}$), and α is the FWHM of the Gaussian kernel ($\alpha = 10\ \mu\text{as}$). Using (2.21) with $d_{\text{true}} = 44\ \mu\text{as}$ gives $d_{\text{blur}} \approx 43\ \mu\text{as}$. If we also consider the finite resolution of the EHT array ($\sim 20\ \mu\text{as}$) this is further decreased to $\approx 42\ \mu\text{as}$. Therefore, we expect both VIDA and REx to recover a diameter of $42\ \mu\text{as}$. Additionally, the slash strength is also modified by the convolution. Fitting the crescent with the GGR template we find $s = 0.32$, which is the value we will take as the ground truth below.

2.3.2 Step 2: Creating Simulated EHT Observations

While VIDA could be applied to the ground-truth images shown in the left column of [Figure 2.3](#), this is not applicable to what the EHT observes. The EHT is a very-long-baseline interferometer and instead observes complex visibilities, $V(u, v)$ which are related to the on sky image through the van Cittert–Zernike theorem ([Thompson et al., 2017](#)):

$$V(u, v) = \int e^{2\pi i(u\alpha+v\beta)} I(\alpha, \beta) d\alpha d\beta. \quad (2.22)$$

In addition, atmospheric and telescope effects can further corrupt the signal. To model these corruption effects we use the `eht-imaging` package ([Chael et al., 2016a, 2018a](#)) to generate realistic simulated data.

2.3.3 Step 3: Generating an Ensemble of Reconstructed Images from Simulated VLBI Data

To validate VIDA we used image reconstructions from the forward modeling or “regularized maximum likelihood methods” (RML), e.g. [Honma et al. \(2014\)](#); [Bouman et al. \(2016\)](#);

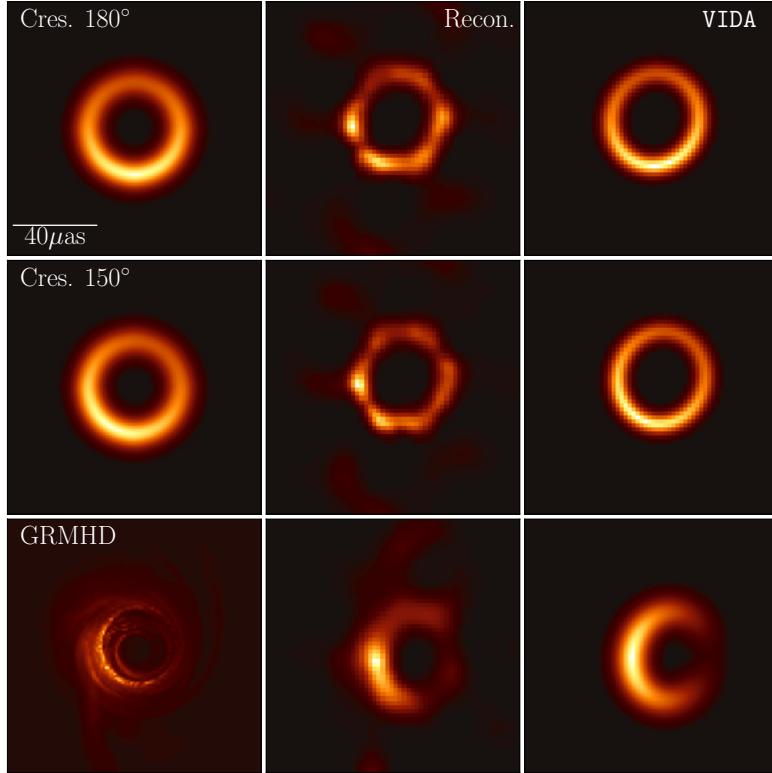


Figure 2.3: Images used for the imaging validation from [EHTC IV](#). We considered 3 models: *top row* crescent with position angle $\xi = 180^\circ$ north of east, *middle row* crescent with position angle $\xi = 150^\circ$ and *bottom row* GRMHD simulation. The left column shows the truth image, the middle columns an example reconstruction, and the right the optimal VIDA template applied to the reconstruction.

[Akiyama et al. \(2017a,b\)](#); [Ikeda et al. \(2016\)](#); [Kuramochi et al. \(2018\)](#) and more specifically the `eht-imaging` package ([Chael et al., 2016a, 2018a](#)). The goal of RML methods is to find the image, I , that minimizes the objective function

$$J(I) = \sum_{\text{data}} \alpha_d \chi_d^2(I) - \sum_{\text{regularizers}} \beta_r S_r(I). \quad (2.23)$$

Following [EHTC IV](#), each χ_d^2 is defined solely from the data products from the EHT telescope, e.g., complex visibilities. The second term encapsulates our additional assumptions,

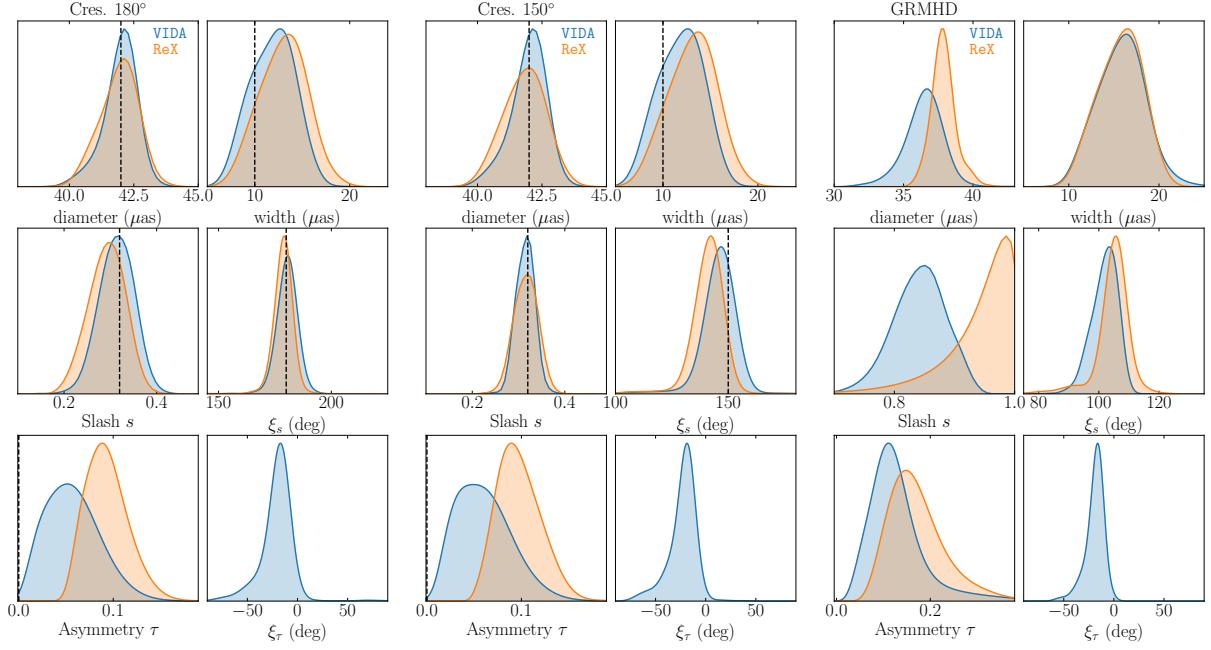


Figure 2.4: Results from **VIDA** (blue) and **REx** (orange) being applied to the simulated two crescent and GRMHD data using the coverage from April 11. The left-two and middle-two columns show the geometric crescent models’ results and the right-two for the GRMHD. In all instances, we found excellent agreement between **VIDA** (blue) and **REx** (orange) for the ring diameter, width, ellipticity, and location of the azimuthal brightness. Furthermore, the crescent distributions for the diameter (after accounting for [Equation 2.21](#)), width, slash strength s and position angle ξ_s are consistent with the truth image (black dashed line). For the ellipticity, τ , both **REx**, and **VIDA** give similar results but are biased from the truth. The origin of this bias will be explored in future work. The ellipticity orientation, ξ_τ , is only recovered by **VIDA**, so there is no **REx** comparison. Similar results were found using the coverage from April 5, 6, and 10.

or regularizers, that are placed on the image. The α_d , β_r , are the “hyperparameters” that control the relative weighting of the regularizers and data products. For the list of the regularizers used, see [EHTC IV](#). In an attempt to model the uncertainty in the image reconstructions, we used the same set of imaging hyperparameters as in [EHTC IV](#). The resulting set of image reconstructions is called the “top set” and results in 1572 reconstructions per dataset.

2.3.4 Step 4: Applying VIDA to the Ring-Like Image Reconstruction Ensembles

VIDA was run on each set of image reconstruction ensembles using the GGR template with a constant flux background whose relative intensity was free parameter, giving 9 parameters in total. Some example reconstructions and corresponding optimal templates are shown in [Figure 2.3](#). The results are shown in [Figure 2.4](#). For the crescent models, we were able to recover the expected diameter d , width w , and azimuthal orientation ξ_s (black dotted lines in [Figure 2.4](#)). In addition, we compared the VIDA results to the method used [EHTC IV](#), [REx](#).

[REx](#) assumes that a single ring-like feature dominates the image reconstruction and then finds the ring by finding the image location that leads to a ring with minimal radial dispersion. [REx](#) characterizes (see [Section 2.2.4](#) and [EHTC IV](#) for definitions) the ring through a diameter d , width w , brightness moments s and orientation ξ_s and a fractional dispersion of the diameter f_d . The diameter and width, and brightness profile of the [REx](#) measurement are similar to VIDA’s measurement with the GGR template. However, the fractional dispersion is not directly measured. Instead, VIDA measures the elliptical ellipticity of the ring. As a result, we expect that the ellipticity measured by [REx](#) will be systematically larger than the one measured by VIDA. More specifically, as is shown in [Figure 2.2](#), we expect [REx](#)’s ellipticity measurement will have a floor of $\tau \sim 0.05$.

The agreement between [REx](#) and VIDA is excellent. The peak and overall width of the distribution for each parameter in [Figure 2.4](#) are consistent between [REx](#) and VIDA. The ground truth values (black vertical lines) for the diameter, width, and brightness orientation ξ_s are also consistent with the [REx](#) and VIDA results.

To compare VIDA and [REx](#)’s measurement of ellipticity, we first note that an additional processing step is needed since the two definitions differ. [REx](#) doesn’t directly measure τ but instead measures the fractional diameter dispersion of the ring f_d (see [Equation 2.14](#) for a definition). If we assume that the ring’s ellipticity dominates f_d , then f_d and τ are related by an invertible map. For more information about this conversion, see [Section 2.2.4](#). In [Figure 2.4](#) we show [REx](#)’s results after converting from f_d to τ . Comparing the two measurements of τ , we see that [REx](#)’s measurement is consistently greater than VIDA’s. This bias is not unexpected given that when τ is small, the conversion described in [Section 2.2.4](#) no longer applies. Instead, the fractional dispersion is dominated by random fluctuations in the ring diameter, creating a floor in f_d . If we then naively apply the previous conversion, as was done in [Figure 2.4](#), we will overestimate τ (see [Figure 2.2](#)).

VIDA also recovers the orientation of the ring ellipticity ξ_τ . Interestingly, in all instances,

we measure a similar distribution for ξ_r irrespective of the intrinsic image. This distribution is biased so that the semi-major axis of the ellipticity is in the north-south direction. By visually inspecting the image reconstruction ensembles (see [Figure 2.3](#) for a typical example reconstruction) we confirmed this feature was present in the reconstructions and was not a bias from VIDA. The origin of this bias will be explored in [Chapter 3](#), where we show this is an inherent weakness of the `eht-imaging` top set.

2.4 Applying VIDA to Additional Simulated Image Reconstructions

In the previous section, we saw that VIDA and REx gave similar answers to problems that demonstrated similar ring-like structures. While REx is limited to ring extraction, VIDA can be applied to any image given a suitable template function. This section will explore VIDA’s capabilities of extracting features from a broader range of potential sources. To accomplish this, we will consider the other non-ring test images from [EHTC IV](#): the symmetric disk and double Gaussian (see [Figure 2.5](#)). We will follow the same steps in the previous section to evaluate VIDA’s performance.

2.4.1 Double Gaussian

Here we consider a source composed of a compact double with two circular Gaussian components. Each Gaussian has an FWHM of $20\mu\text{as}$. One of the Gaussians is placed at the origin and has a flux of 0.27Jy , which we will call the NW component. The other Gaussian is at $\Delta\text{RA} = 30\mu\text{as}$ and $\Delta\text{DEC} = -12\mu\text{as}$ and has a flux of 0.33Jy and will be called the SE component. This type of source could arise when looking at AGN using VLBI images, such as the recent 3c 279 results ([Kim et al., 2020](#)).

To extract the reconstruction’s compact components, we used a template with three asymmetric Gaussian components and constant background. Two of the Gaussian components were allowed to be arbitrary, while the third Gaussian component was forced to be large ($r_0 > 15\mu\text{as}$). The reason for the third Gaussian was there tended to be a region of additional flux around the two dominant Gaussian components in the image reconstructions. This diffuse flux can be seen in the top middle panel of [Figure 2.5](#). If we didn’t include this third component, we found that the Gaussian components tended to be quite large to soak up the extra flux.

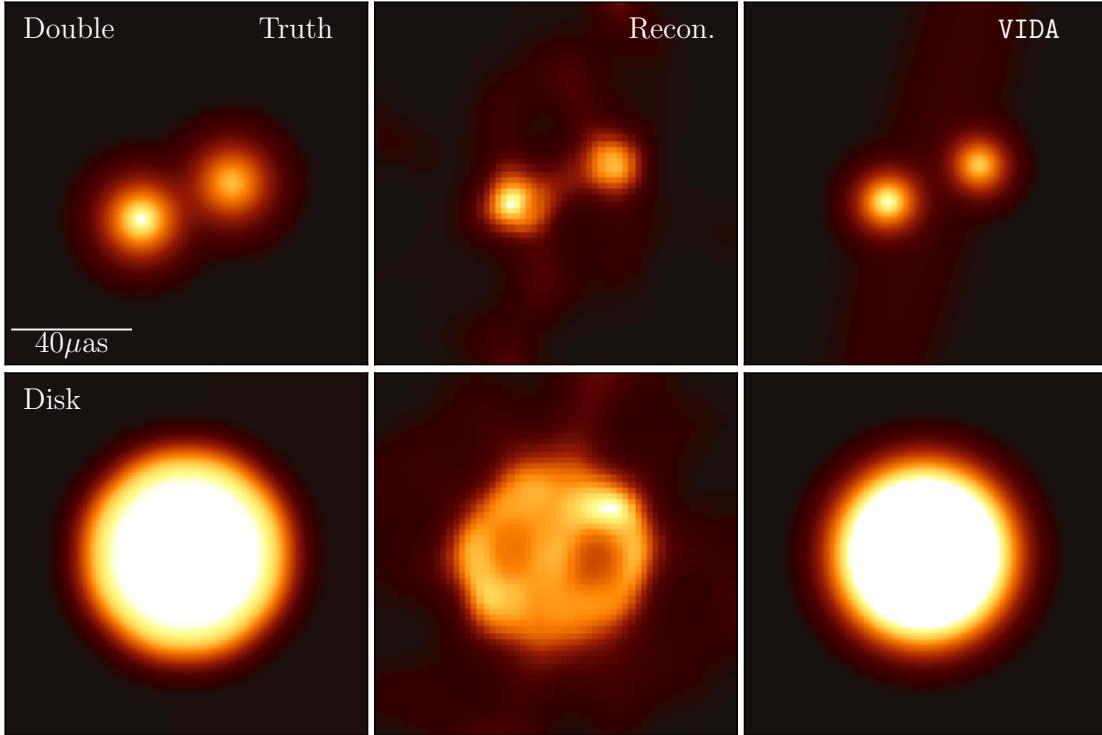


Figure 2.5: Results from applying VIDA to the non-ring test images from EHTC IV where the left column shows the ground truth image, the middle an example reconstruction from the top set, and right the optimal template from applying VIDA to the reconstruction. *Top:* Shows the results for the double image. *Bottom:* Shows the same results for the disk.

VIDA's results for the double Gaussian are shown in Figure 2.6. Overall the size of each Gaussian, their separation, and flux ratio are reliably recovered. The ellipticity, τ , is larger than zero, but this is to be expected since the algorithm can only add ellipticity to the Gaussian components.

On April 6, we see that the ellipticity appears to be bimodal, and the parameter uncertainties are greater than the other days. This uncertainty was unexpected, given the relatively good EHT coverage on April 6. The origin of the discrepancy is from a subset (10 – 15%) of image reconstructions that exhibit an additional bright Gaussian feature. If we remove these reconstructions, we find that the results on April 6 are consistent with the other days.

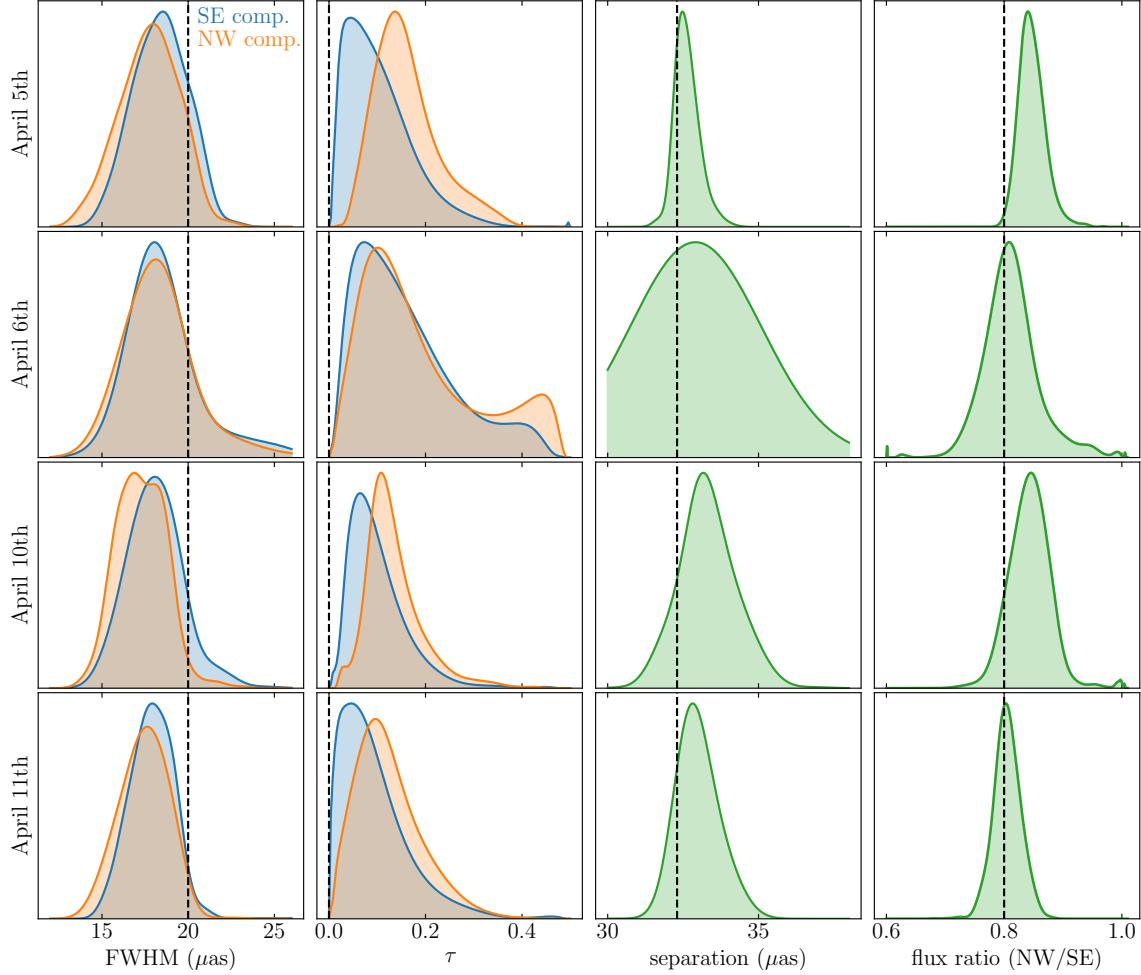


Figure 2.6: VIDA results for the two compact Gaussian components (blue for SE component and orange for the NW) in the double Gaussian test image. The green curves are for parameters that are a combination of the SE and NW components. On all days the true values are included in the parameter distributions found by VIDA. Note, that the broad distribution found on April 6th is due to an imaging artifact as discussed in the chapter.

2.4.2 Disk Image

The intrinsic image is a symmetric flat disk with a diameter of $70 \mu\text{as}$, which is then convolved with a Gaussian with an FWHM of $10 \mu\text{as}$. The true image, an example recon-

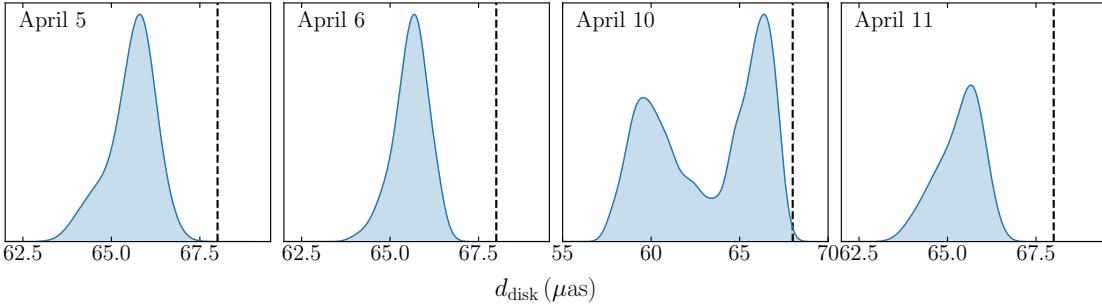


Figure 2.7: VIDA results of the diameter for the disk top set. The diameter is given by (2.24). Ignoring April 10, we consistently find that the diameter is $4\mu\text{as}$ smaller than the original image. The origin of this discrepancy is discussed in Figure 2.8. On April 10, which has poor coverage compared to the other days, the imaging gives two modes. One mode is similar to the other days, while the second fails to show a coherent disk structure giving the second peak at $60\mu\text{as}$.

struction, and optimal template for that reconstruction are shown in the lower left, middle panel, and right panel of Figure 2.5.

To encode the diameter of the disk we use FWHM of the disk template:

$$d_{\text{temp}} = 2r_0 + 2\sqrt{2 \log 2}\alpha, \quad (2.24)$$

where r_0 and α are described in Equation 2.1. We fit the disk template to the ground truth image to calibrate the diameter definition to the disk’s true diameter. We found that the optimal template for the true image had a $d_{\text{temp}} \approx 69\mu\text{as}$. If we convolved the image by an additional $20\mu\text{as}$ to take into account the finite resolution of the EHT array, we found $d_{\text{temp}} \approx 68\mu\text{as}$. This is the value we use as the ground truth diameter in all comparisons below.

Figure 2.7 displays the results for VIDA applied to each day. Ignoring April 10, which has poor coverage compared to the other days, we find that the results are very consistent between days. On April 5, we find the median diameter $d_{\text{temp}} = 65.7^{+0.42}_{-0.76}\mu\text{as}$ where the range are the 68% interval about the median. Similarly, on April 6th we find $d_{\text{temp}} = 65.6^{+0.40}_{-0.47}\mu\text{as}$, and April 11th $d_{\text{temp}} = 65.5^{+0.45}_{-0.80}\mu\text{as}$. This demonstrates that VIDA is robust to the slight difference in image reconstructions from different baseline coverages.

On April 10, however, we had a different result finding a bimodal diameter. Analyzing the reason for this, we found that images with $d_{\text{disk}} \approx 60\mu\text{as}$ had a markedly different structure than the rest of the images. Given the distinct non-disk structure of the image, it is no surprise that VIDA struggles at recovering the correct diameter.

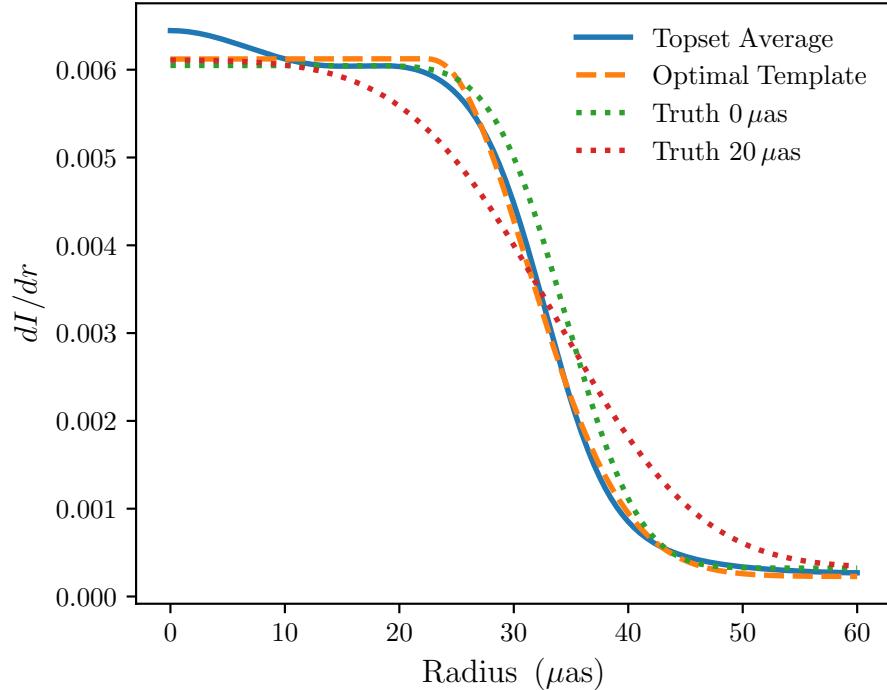


Figure 2.8: Results when applying VIDA to the average disk top set reconstruction on April 11. The images were normalized to unit flux and centered before averaging. The average radial profile is shown in solid blue. Comparing this to the optimal template (orange dashed line) and the true profile (dotted lines), we see that the optimal template matches the reconstruction’s radial profile but underestimates the ground truth image.

Comparing our result for the diameter to the true value, $68\mu\text{as}$, we find that our result has a consistent bias of $\approx 2.4\mu\text{as}$ on April 5, 6, and 11. Again, this appears to be an artifact of the imaging process. In Figure 2.8, the radial profiles of the truth (dotted lines), averaged image reconstructions⁶, and optimal templates are compared. VIDA does an excellent job of recovering the size of the images, which are similarly biased toward smaller radii. This suggests that the diameter bias is intrinsic to the top set used for this disk. As discussed above, it is unlikely that this is due to the finite resolution of the EHT array and is intrinsic to the imaging algorithms used.

⁶When averaging we first centered the images by computing the image centroid and normalized the images to have unit flux.

2.5 Summary and Conclusions

We present VIDA, a new image feature extraction technique appropriate for use by the EHT. VIDA adopts a forward modeling approach to extract quantitative image properties by approximating the image with a parameterized family of functions that encode the desired image properties.

A key feature of VIDA is its flexibility. Multiple image components have already been implemented, from which composite models of significant complexity can be constructed. These include ring-like templates of particular relevance to EHT images, Gaussians, and diffuse backgrounds.

The ability of VIDA has been demonstrated for several sources, each with over a thousand reconstructions. These include image reconstructions from simulated data produced from double Gaussians, slashed rings, and GRMHD simulations. In all cases, key quantitative features were accurately recovered where they appeared in the underlying image reconstructions. These include separations, orientations, ring diameters, widths, brightness profiles, and multiple measures of ellipticity. Application of these to the EHT observations of M 87 will be explored in the next chapter.

The applicability of VIDA extends beyond EHT observations of M 87. The ability to create composite models with multiple components is naturally relevant to VLBI reconstruction of AGN, such as 3C 279, that is composed of multiple compact features (e.g., Kim et al., 2020).

It should be noted that image feature extraction methods, like VIDA, are generally most useful when strong priors may be placed on the image structure itself. That is, VIDA is primarily a method for quantifying what is already qualitatively apparent. Poorly chosen models can lead to significant parameters biases, as seen in Section 2.4.1, where an extra Gaussian blob was required to achieve acceptable results. However, because VIDA is an image characterization tool, not an imaging tool in and of itself, this presents only a very modest limitation on its utility.

In the next chapter we will apply VIDA to the `eht-imaging` reconstructions of M 87. Furthermore, since VIDA can recover ring ellipticity, through the τ parameter, we will be able to measure the ellipticity of M 87. We will also investigate the origin of the $\xi_\tau = 0$ bias seen in the geometric reconstruction in Figure 2.4.

Chapter 3

Constraining the ellipticity of M 87 from EHT Image Reconstructions

3.1 Introduction

The Event Horizon Telescope (EHT) can resolve the emission around the event horizon of the supermassive black hole M 87 ([EHTC I](#); [EHTC II](#); [EHTC III](#); [EHTC IV](#); [EHTC V](#); [EHTC VI](#)). Using novel imaging techniques developed by the EHT collaboration, a ring-like emission structure was observed. The measured ring radius was consistent with a central black hole with mass $6.5 \times 10^9 M_\odot$. In addition to the mass, of particular interest for gravitational physics is the structural ellipticity, or ellipticity, of the ring. If the no-hair theorem breaks down near the event horizon of supermassive black holes, the shadow may be deformed, inducing additional non-circularity (see, e.g., [Johannsen & Psaltis, 2010](#); [Broderick et al., 2014](#); [Johannsen et al., 2016](#); [Medeiros et al., 2020](#)). Therefore, using the EHT image reconstructions to measure the image ellipticity may constrain near-horizon deviations from GR.

In [EHTC VI](#), a preliminary attempt at measuring the ellipticity of the image reconstructions of M 87 was presented in Figure 18. To measure the ellipticity, they used REx’s fractional dispersion detailed in the previous chapter. The measured dispersion implied an axis ratio of 4 : 3, suggesting that M 87 images were highly symmetric. However, the interpretation of this measurement is not straightforward.

While the EHT can resolve the horizon scale structure, its dynamic range and visibility coverage is poor ([EHTC II](#); [EHTC III](#)). As a result, infinitely many images can reproduce

the data, and additional assumptions are imposed to make the problem tractable. In [EHTC IV](#), regularized maximum likelihood (RML) methods were applied to the M 87 data. RML introduces the additional assumptions through regularizers that enforce features such as image smoothness ([Bouman et al., 2016; Chael et al., 2016a; Kuramochi et al., 2018](#)), sparseness ([Wiaux et al., 2009a,b; Honma et al., 2014; Akiyama et al., 2017b](#)), and similarity to some fiducial image ([Narayan & Nityananda, 1986](#)). However, the weights of these regularizers are unknown. Therefore, the infinite image problem has been shifted to selecting the weights of the regularizers. To combat this [EHTC IV](#), performed a parameter survey using simulated data and selected parameters that met some specified criteria. The parameter survey used in [EHTC IV](#) did not include an asymmetric ring in simulated tests, and thus the reliability of the [EHTC IV](#) ellipticity measurement is unclear. Furthermore, these image surveys do not form a posterior, and any ellipticity recovered may be biased. Whether the imaging parameter survey used in [EHTC IV](#) can reliably recover ellipticity is, therefore, an open question.

In the previous chapter, we presented **VIDA**, a novel feature extraction technique for the EHT image reconstructions. In this chapter, we will apply **VIDA** to the imaging parameter surveys of M 87 to assess their ellipticity reconstruction fidelity. To evaluate whether the surveys can reliably recover image ellipticity, we consider a new set of geometric tests specifically targeted to measure ellipticity. We will show that the imaging surveys are not able to reliably recover the ellipticity and its orientation. To overcome these issues, we will then run the imaging surveys on set of GRMHD simulations to calibrate for the imaging ellipticity bias. This is similar in spirit to mass calibration procedure done in [EHTC VI](#), and is necessary to interpret the ellipticity results from the imaging pipelines.

The layout of the chapter is as follows: In [Section 3.2](#) we will review the M 87 imaging survey, feature extraction techniques and the M 87 ellipticity results. Next in [Section 3.3](#), we show that the current parameter survey used in M 87 is not able to recover ellipticity. Finally, we will analyze simulated data from 550 GRMHD simulations to calibrate the image reconstruction ellipticity of the M 87 results.

3.2 Background

This section will review the standard image reconstruction techniques used by the EHT in [EHTC IV](#). The imaging techniques used in this thesis will be identical to the `eht-imaging` pipeline used in [EHTC IV](#). After, we will review the two feature extraction techniques we use, i.e., **REx** and **VIDA**. Finally, we will apply **REx** and **VIDA** to the M 87 top set. As a

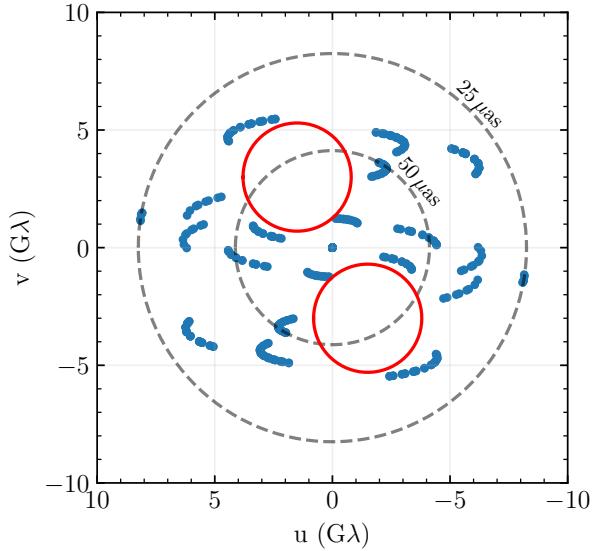


Figure 3.1: UV coverage of the EHT 2017 array on April 11. The blue dots denote the 2017 EHT coverage on April 11 in units of $G\lambda$. The black dotted lines show the characteristic locations for the image features of size $50 \mu\text{as}$ and $25 \mu\text{as}$ in the uv domain. The red circles highlight the coverage gap in the north-south direction.

result, we will reproduce the results from [EHTC VI](#) and extend the analysis to include the orientation of M 87’s ellipticity.

3.2.1 Image Reconstructions and the M 87 top set

Our goal is to assess the capabilities of the parameter surveys, aka top sets, used for M 87. Therefore, we will use the results outlined in [Section 2.3](#). Namely, the M 87 top set is found by fitting the simulated data sets including symmetric, crescents, rings, double Gaussians, disks, and a small number of GRMHD snapshots. This top set will be applied to all simulated data sets, including the elliptical ring models in [Section 3.3](#) and GRMHD sets in [Section 3.4](#).

3.2.2 Feature Extraction Techniques

Variational Image Domain Analysis VIDA

Recall that variational image domain analysis (VIDA) requires three ingredients: an image I , a template image f_θ , and a probability divergence \mathcal{D} that measures the difference between f and I . VIDA relies on the template function f_θ being a reasonable approximation to the true image. Given that we are interested in ring morphologies, we will use the `CosineRing{N, M}` template¹. Section 3.3 is interested in recovering the profile for a simple elliptical ring with a slash. Therefore, we will take $N = 0$ and $M = 1$. We also tried higher-order mode expansions and found that they were typically much smaller than the first mode and did not change the results.

For the GRMHD reconstructions in Section 3.4 we take $N = 1, M = 4$, given their complicated azimuthal structure. This template has 16 parameters in total. To characterize the thickness or slash strength of the template, we will refer to the σ_0 and s_1 parameters respectively. In addition to the above templates, we add a constant flux background where the background intensity is also a parameter. This models the diffuse flux that is typically found in image reconstructions.

To find the optimal template we will use the Bh divergence (Equation 2.8) from the last chapter. To minimize the Bh divergence we use the same optimization strategy that was validated in Section 2.3.

Ring Extractor REx

The other image feature extraction method used in this thesis is the *ring-extractor* or REx algorithm used in EHTC IV and described in detail in Chael (2019). REx's parameter definitions and relation to VIDA's parameters are detailed in Section 2.2.4. Note that we automatically convert REx's fractional dispersion to τ for all discussion and figures below.

3.2.3 Review of M 87 Ellipticity Measurement

Figure 18 of EHTC VI, showed the measured M 87 fractional deviation from REx applied to the eht-imaging top set parameters. To reproduce these results, we applied VIDA and

¹For more information about the other templates present in VIDA, please see <https://github.com/ptiede/VIDA.jl> and Section 2.2.1

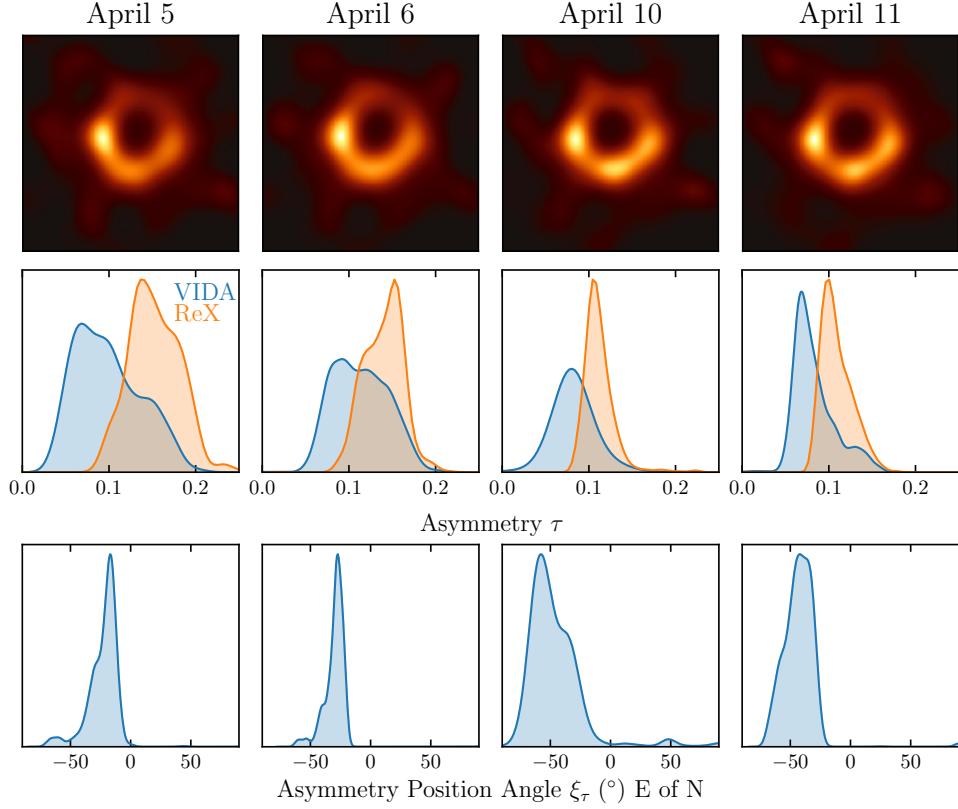


Figure 3.2: `eht-imaging` reconstructions and ellipticity features of M 87 from the [EHTC IV](#) top set across the different observations during the 2017 EHT campaign. The top row shows the fiducial `eht-imaging` images from [EHTC IV](#). The middle row shows the measured ellipticity of `eht-imaging` top set from VIDA (blue) and REx (orange). The bottom row shows the orientation of the measured ellipticity orientation angle east of north from VIDA. No ellipticity orientation angle is shown for REx, since it is not currently able to measure it.

REx to the `eht-imaging` top set. The results are shown in Figure 3.2. The top row shows the fiducial image reconstruction from the `eht-imaging` top set across each observation day. The ellipticity of the top set images is shown in the middle row. We find identical results to those in [EHTC VI](#) for REx. The VIDA results are systematically lower than REx, as expected from the discussion in [Section 2.2.4](#). The bottom row presents, for the first time, the orientation of this ellipticity from VIDA. Note that REx cannot currently measure this orientation. Overall we find the ellipticity measurement is stable across all four days,

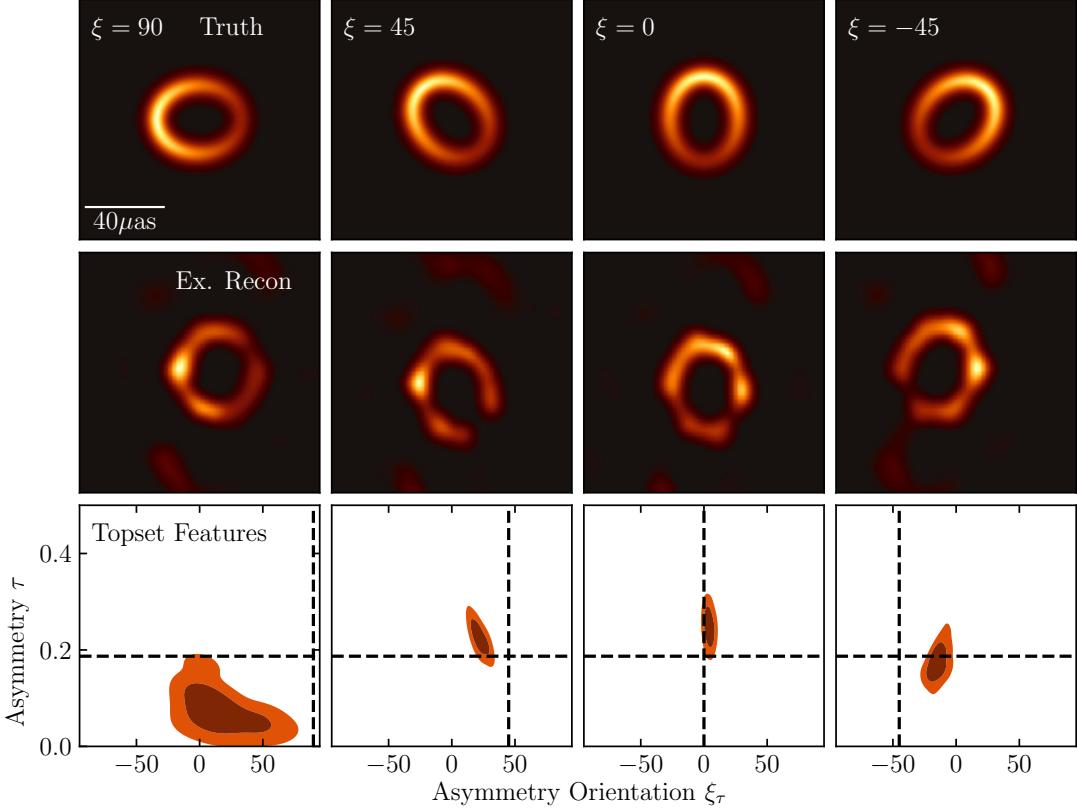


Figure 3.3: Examples of image reconstructions and VIDA results for the different elliptical rings. The top row shows the truth images at position angles $\xi_s = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ north of east. The middle row shows an example reconstruction from the M 87 top set for each ring orientation. The bottom row shows the VIDA results for the ellipticity and its positions angle from the top set with the 68% and 95% probability regions shown. We found significant bias in τ and ξ_τ for rings whose semi-major axis was aligned in the east west direction.

giving $\tau = 0.05 - 0.2$, and orientation $\xi_\tau = -75^\circ - 0^\circ$ east of north.

While it is interesting that the ellipticity measurements are consistent across days, it is not clear whether this result is intrinsic to the source. In [Chapter 2](#), we found statistically identical ellipticity and orientation for the symmetric crescent models and GRMHD models. Given that the crescent models are symmetric, it suggests that the measured ellipticity may be an imaging artifact. Furthermore, the ellipticity orientation does align with a coverage

Image			d_0	w	τ	$\xi_\tau(\circ)$	s	$\xi_s(\circ)$
$\xi_s = 90^\circ$	$\xi_\tau = 90^\circ$	REx	$37.3_{-1.0}^{+1.2}$	$12.4_{-4.4}^{+4.4}$	$0.09_{-0.04}^{+0.07}$...	$0.54_{-0.09}^{+0.07}$	$90.68_{-7.8}^{+3.9}$
		VIDA	$37.5_{-1.2}^{+1.3}$	$11.1_{-4.0}^{+3.5}$	$0.07_{-0.05}^{+0.09}$	$12.6_{-22.8}^{+48.8}$	$0.54_{-0.07}^{+0.05}$	$91.4_{-10.4}^{+3.9}$
$\xi_s = 45^\circ$	$\xi_\tau = 45^\circ$	REx	$37.2_{-1.2}^{+0.8}$	$11.2_{-3.6}^{+4.7}$	$0.20_{-0.04}^{+0.04}$...	$0.52_{-0.12}^{+0.11}$	$48.9_{-7.8}^{+5.3}$
		VIDA	$37.7_{-1.4}^{+0.9}$	$10.0_{-3.4}^{+3.9}$	$0.23_{-0.03}^{+0.04}$	$21.9_{-7.2}^{+8.1}$	$0.52_{-0.12}^{+0.18}$	$48.0_{-9.8}^{+6.5}$
$\xi_s = 0^\circ$	$\xi_\tau = 0^\circ$	REx	$37.0_{-1.3}^{+0.9}$	$11.2_{-3.7}^{+4.3}$	$0.24_{-0.03}^{+0.05}$...	$0.45_{-0.07}^{+0.15}$	$-5.7_{-20.6}^{+33.3}$
		VIDA	$37.5_{-1.3}^{+0.7}$	$9.8_{-3.3}^{+3.8}$	$0.25_{-0.04}^{+0.04}$	$3.4_{-3.6}^{+5.3}$	$0.45_{-0.07}^{+0.13}$	$-3.4_{-18.0}^{+27.1}$
$\xi_s = -45^\circ$	$\xi_\tau = -45^\circ$	REx	$37.5_{-1.1}^{+1.1}$	$11.7_{-3.9}^{+4.3}$	$0.18_{-0.04}^{+0.04}$...	$0.52_{-0.09}^{+0.09}$	$-49.4_{-5.6}^{+22.1}$
		VIDA	$38.3_{-1.3}^{+0.9}$	$10.5_{-3.5}^{+3.5}$	$0.17_{-0.05}^{+0.06}$	$-15.8_{-9.2}^{+8.8}$	$0.53_{-0.04}^{+0.13}$	$-50.7_{-6.0}^{+19.8}$
$\xi_s = -90^\circ$	$\xi_\tau = 90^\circ$	REx	$37.5_{-1.2}^{+1.1}$	$12.4_{-4.4}^{+4.4}$	$0.09_{-0.05}^{+0.07}$...	$0.54_{-0.12}^{+0.07}$	$-90.3_{-6.6}^{+6.4}$
		VIDA	$37.8_{-1.2}^{+1.2}$	$11.1_{-3.9}^{+3.5}$	$0.06_{-0.05}^{+0.09}$	$12.6_{-18.6}^{+42.4}$	$0.54_{-0.08}^{+0.06}$	$-90.0_{-8.6}^{+6.2}$
$\xi = -135^\circ$	$\xi_\tau = 45^\circ$	REx	$37.1_{-1.2}^{+0.7}$	$11.2_{-3.6}^{+4.6}$	$0.20_{-0.03}^{+0.15}$...	$0.47_{-0.09}^{+0.12}$	$-131.4_{-6.0}^{+5.4}$
		VIDA	$37.6_{-1.2}^{+0.8}$	$9.9_{-3.5}^{+3.8}$	$0.22_{-0.03}^{+0.04}$	$21.3_{-7.1}^{+8.8}$	$0.52_{-0.09}^{+0.19}$	$-134.4_{-7.7}^{+4.4}$
$\xi = -180^\circ$	$\xi_\tau = 0^\circ$	REx	$37.0_{-1.3}^{+0.9}$	$11.1_{-3.7}^{+4.2}$	$0.24_{-0.04}^{+0.04}$...	$0.47_{-0.08}^{+0.16}$	$-184.1.2_{-27.0}^{+20.4}$
		VIDA	$37.6_{-1.2}^{+0.8}$	$9.8_{-3.3}^{+3.9}$	$0.25_{-0.03}^{+0.03}$	$4.3_{-3.5}^{+3.7}$	$0.49_{-0.07}^{+0.16}$	$-183.1_{-12.9}^{+32.6}$
$\xi = 135^\circ$	$\xi_\tau = -45^\circ$	REx	$37.6_{-1.1}^{+1.2}$	$11.9_{-4.3}^{+4.0}$	$0.18_{-0.03}^{+0.04}$...	$0.53_{-0.05}^{+0.08}$	$127.9_{-5.3}^{+8.3}$
		VIDA	$38.3_{-1.2}^{+0.8}$	$10.5_{-3.3}^{+3.9}$	$0.17_{-0.03}^{+0.03}$	$-13.2_{-3.5}^{+3.7}$	$0.56_{-0.07}^{+0.16}$	$127.9_{-6.1}^{+7.5}$
Truth			37.56	7.9	0.187	...	0.5	...

Table 3.1: Recovered parameters for the slashed elliptical rings test set. The parameters are the median values and the 95% interval around the median. REx and VIDA give similar results for all parameters, although no results for ξ_τ are given for REx since it cannot recover it. All recovered parameter ranges, except the ellipticity τ and its orientation, ξ_τ contain the true values.

gap (see the red circles in Figure 3.1). Finally, since the top set used for M 87 did not include an ellipticity ring, it is not clear whether ellipticity can reliably be recovered. To investigate fidelity of ellipticity reconstruction using the top set, we will apply it to simple asymmetric ring images in the next section.

3.3 Geometric Test

One of the potential issues with the M 87 top set is that no asymmetric rings were included in the simulated data tests. Given that the identification of the top set was defined by its performance on simulated data tests, the top set may not accurately recover ring ellipticity even in simple cases. This section will analyze simulated data from an asymmetric ring model using the M 87 top set.

3.3.1 Asymmetric Image Test

For the asymmetric image we used the `CosineRing{0,1}` template described in [Section 3.2.2](#), with parameters $d_0 = 37.56 \mu\text{as}$, $w = 7.9 \mu\text{as}$, $\tau = 0.187$, $s = 0.5$. The ring flux was set 0.6 Jy, matching the measured compact flux of M 87. We also aligned the orientation of the slash and ellipse, i.e., we set $\xi_s = \xi_\tau = \xi$. To test the impact of different orientations of the ellipticity we considered $\xi = 0^\circ$ to 360° in steps of 45° . A subset of the ground truth images are shown in the top row of [Figure 3.3](#). For each rotated ring we used the `eht-imaging` top set pipeline from [Section 2.3.3](#), to create 1572 reconstructions.

3.3.2 Geometric Results

Given the elliptical ring reconstructions, we then used VIDA and REx to extract the relevant image features. Since the VIDA filter is identical to on-sky image, we would expect to perfectly recover every parameter. The results for each orientation are summarized in [Table 3.1](#). [Table 3.1](#) shows that the ring diameter and width are consistent across the rotation angles, and are consistent with the truth. The slash and its orientation are similarly recovered.

However, the ellipticity, τ , is significantly biased when $\xi_\tau = 90^\circ$, i.e., when the semi-major axis of the ellipse is aligned in the east-west direction. Furthermore, looking at the bottom row of [Figure 3.3](#), we see that the orientation of the ellipticity is consistently biased towards $\xi_\tau = 0$, i.e., the north-south direction. This bias can be visually confirmed by looking at the reconstructions, e.g., the middle row of [Figure 3.3](#). Furthermore, we see that the true ellipticity and orientation is only recovered in the $\xi = 0$ case.

Therefore, it appears that imaging creates a preferred ellipticity direction $\xi_\tau \approx 0$, τ . Namely, as τ increases, the ellipticity orientation tends to point in the north-south direction. This orientation does approximately align with a large gap in the EHT coverage for M 87 (see [Figure 3.1](#)).

Given that the top set fails to recover ellipticity in our geometric tests, the ellipticity measurement found in [Figure 3.2](#) does not represent the true ellipticity of the on-sky image. Instead, it measures an artifact from the top set. To account for this artifact, we need to calibrate our result. In the next section, we will image a large number of GRMHD simulations to measure the imaging ellipticity bias and uncertainty.

3.4 Calibrating the M 87 Ellipticity Measurement

To do calibrate the M 87 topset ellipticity we will use a similar procedure to the mass calibration done in [EHTC VI](#). Namely, we will select a number of GRMHD simulations from the [EHTC V](#) library, and image them using the `eht-imaging` top set. Using this set of reconstructions, we will infer ellipticity uncertainty and analyze what constraints on M 87's ellipticity can be made.

3.4.1 Scaled set

To construct the GRMHD images used in this chapter, we first cut the simulations from [EHTC V](#) based on whether its total jet power was consistent with the observed jet power of M87 (Table 2 [Event Horizon Telescope Collaboration et al., 2019e](#)). After, we randomly selected 100 snapshots from this set and randomly assigned them a 2017 M 87 observation day. Each image was then re-scaled to its best fit value according to the average image scoring results of [Event Horizon Telescope Collaboration et al. \(2019f\)](#), and randomly rotated. To include the effects of the mass uncertainty of M 87, we then further scaled the intrinsic image by a factor of 0.8, 0.9, 1.1, 1.2 in both the x and y directions. The net result is 500 images uniformly sampled over days, orientations and grided in mass relative to the M 87 best fit value. We will refer to this list as the *scaled set*.

3.4.2 Stretched set

While the scaled set measures the expected ellipticity due to imaging and accretion turbulence, it does not measure how sensitive imaging is to additional intrinsic ellipticity that may occur from, e.g., non-GR spacetimes. To assess the ability to measure a non-symmetric shadow, we randomly selected 10 additional GRMHD snapshots that fit M87's jet power. For each image we scaled them to their best fit mass, and randomly rotated

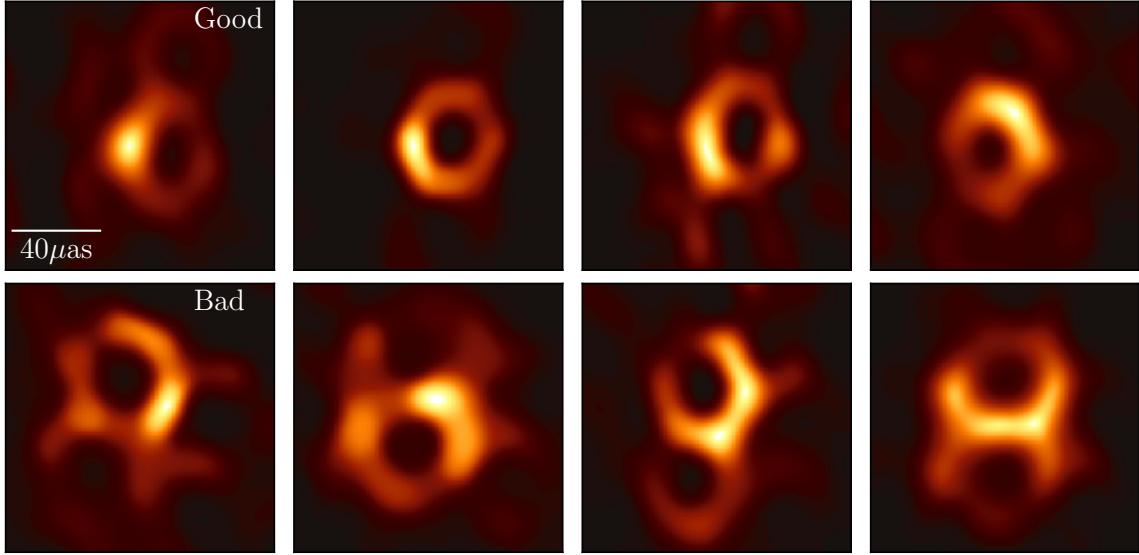


Figure 3.4: Examples of image classification from the trained CNN. The top row shows image reconstructions that passed the threshold, and the bottom are ones that failed. Most of the failed reconstructions fail to demonstrate a single dominant ring-like feature in the image.

them. Ellipticity was then added by picking two random orthogonal directions in the image r_x , r_y and applying the transformation $r_x \rightarrow \alpha r_x$ and $r_y \rightarrow r_y/\alpha$. To create different amounts of ellipticity we let $\alpha = 0.8, 0.9, 1.0, 1.1, 1.2$, giving a $\tau = 0.36, 0.19, 0.0, 0.17, 0.31$ respectively. We will refer to this as the *stretched set*.

3.4.3 Removing failed top set reconstructions

While inspecting the top set reconstructions of the scaled and stretched GRMHD simulations we noticed that a large number of images failed to show a ring-like feature. Instead the image reconstructions had flux deposited across the image in pattern similar to the EHT dirty beam. This is commonly known as “waffling” and is symptomatic of a poorly chosen set of hyperparameters. For these reconstructions, VIDA and REx would give nonsensical results since no dominant ring feature exists. To remove this bias these reconstructions need to be removed. Unfortunately, no single set of hyperparameters was identified as having caused the waffling. Therefore, we turned to machine learning techniques to remove

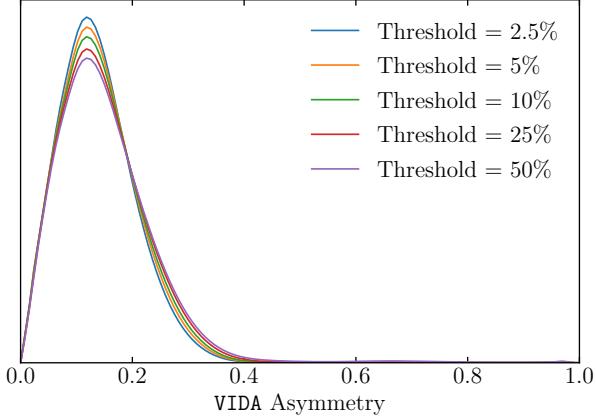


Figure 3.5: Impact of the threshold on the inferred ellipticity of the scaled reconstructions. Overall there is very little change in the distribution if the threshold is ≤ 0.5 . In this thesis we use the threshold of 0.10 which removes 42% of the scaled reconstructions.

any waffled images.

Machine learning and neural networks, have been used numerous times in astrophysical settings to classify images (Davelaar et al., 2017; Dieleman et al., 2015, e.g.). To classify our images we decided to use convolutional neural networks (CNNs) (e.g. Goodfellow et al., 2016). CNN's break the images into features of different scales and then group these features to classify the image. For our neural network we used the Julia package Flux (Innes et al., 2018; Innes, 2018).

For our image classifier we used a relatively shallow network using 3 convolutional layers with a 2×2 max pooling. For the final layer we used a fully connected network to the two-dimensional classification space. Between each convolutional layer we used the ReLu activation function: $f(x) = \max(0, x)$. For the final fully connected layer no activation function was used. Since we are interested in binary classification we used the `logitbinarycrossentropy` in Flux, which is given by:

$$H(q) = - \sum_{i=1}^N y_i \log(\sigma(q)) + (1 - y_i) \log(1 - \sigma(q)), \quad (3.1)$$

where $\sigma(x) = (1+e^{-x})^{-1}$ and y_i are the labels (1 for an image that waffled and 0 otherwise). This choice of the loss function is equivalent to using a sigmoid activation function in the last layer of the neural network but has better numerical stability.

CNN’s are a form of supervised learning. Therefore, we first had to label a subset of the image reconstructions by hand. To find the labels, y_i , we analyzed 5000 random images from the scaled GRMHD set and an additional 500 from the stretched set. We then classified each image by whether it visually had a dominant ring-like feature or not. Some examples of images that passed and failed are shown in [Figure 3.4](#). Two-thirds of these classified images were used for training, and the rest for our test set. To combat overfitting, we also augmented the images by adding Gaussian random noise to each image when evaluating the loss function. Finally, ADAM ([Kingma & Ba, 2014](#)) was used to optimize the network using the options defined in the Flux model zoo package² with some minor changes. Namely, we broke our images into batches, with each batch containing 256 images and used a learning rate of 3×10^{-3} . However, if predictive performance on the training set did not improve after 10 epochs we dropped the learning rate by a factor of 10. The optimizer was run for 100 epochs and achieved an accuracy of 94% and 92% on the training and testing set, respectively.

The trained neural network outputs a number a_{waffle} between 0 and 1, that measures the confidence that the image has waffled. To classify whether an image had waffled any reconstruction with $a_{\text{waffle}} > a_{\text{thresh}} = 0.1$ were cut. This threshold cut 42% of images in the scaled set and 21% of the images in the stretched set. The impact of the value of a_{thresh} on the recovered ellipticity is shown in [Figure 3.5](#). As an additional test of the network, we ran the classifier on the elliptical Gaussian reconstructions. We found that only $\sim 5\%$ of the elliptical images were cut. This results is consistent with our visual inspection.

3.4.4 Scaled set results

To analyze the scaled set of images, we used VIDA’s `CosineRing{1,4}` template and Bh divergence. Using the CNN image classifier [Section 3.4.3](#) we removed the “bad” reconstructions, leaving 454,888 images. To compare the results to the GRMHD simulations ellipticity we first blurred the ground truth snapshots with a Gaussian kernel with FWHM $15\mu\text{as}$ to model the finite resolution of the EHT array. Then we fit the blurred images with the VIDA and the `CosineRing{1,4}` template. This formed our ground-truth ellipticity for which we will compare all results below.

The results for ξ_τ and ellipticity τ for the entire scaled set is shown in the upper left panel of [Figure 3.6](#). We found that the ellipticity is quite uncertain in both VIDA and REx extending to $\tau = 0.3$, which is approximately a 5 : 3 axis ratio. This ellipticity is larger

²<https://github.com/FluxML/model-zoo/blob/master/vision/mnist/conv.jl>

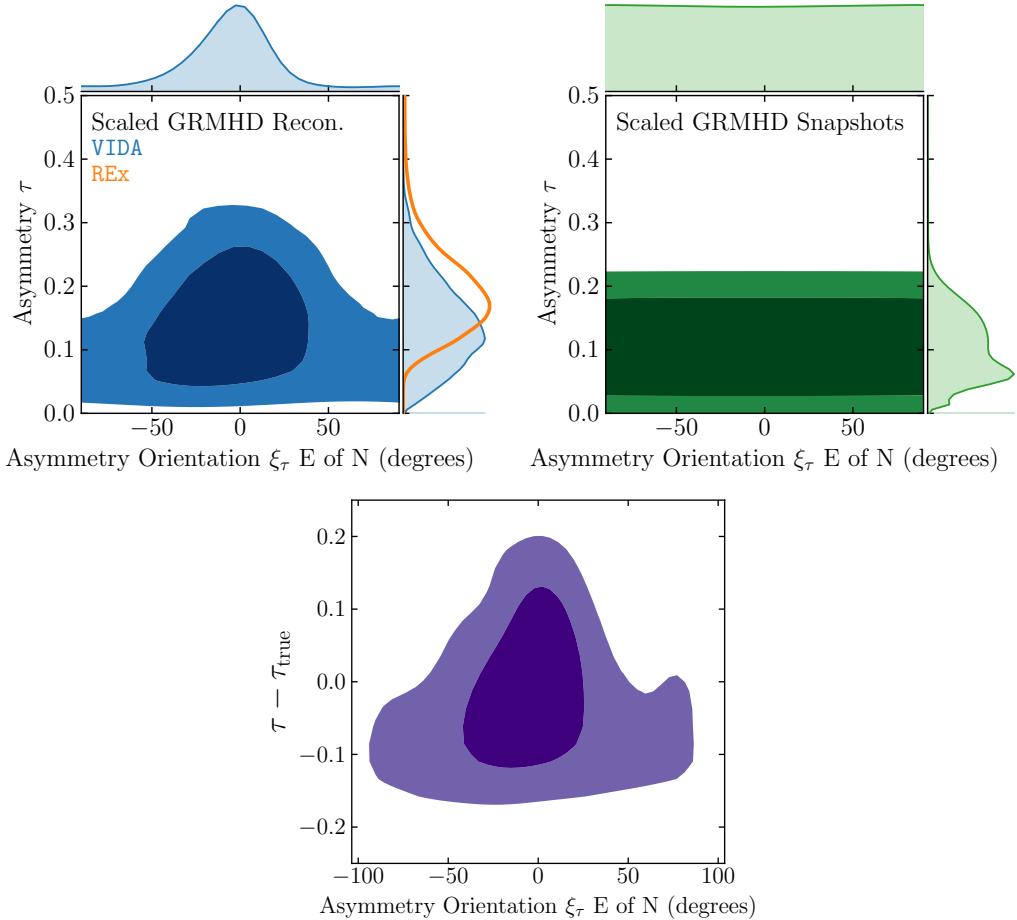


Figure 3.6: Joint marginal probability distribution between τ and ξ_τ , where the contours are the 68% and 95% regions. The upper left shows the results for the entire reconstructed scaled GRMHD set that satisfies $a_{\text{waffle}} < a_{\text{thresh}} = 0.1$ threshold. The upper right is when VIDA is applied directly to the GRMHD snapshot blurred with a $15 \mu\text{as}$ FWHM Gaussian kernel. Like the geometric results, we see a preference for $\xi_\tau \approx 0^\circ$, i.e., the north-south direction regardless of the intrinsic image distribution. The bottom figure shows the measured ellipticity orientation on the x-axis with the measured ellipticity residual distribution.

than ground truth ellipticity measured from VIDA (right panel of Figure 3.6), which is quite concentrated at $\tau \approx 0.1$ and extends up to $\tau = 0.25$.

Focusing on the ellipticity orientation ξ_τ , we find the ellipticity is strongly biased in

the north-south direction. Furthermore, when $\xi_\tau \approx 90^\circ$, τ extends to highest values of ellipticity. This orientation distribution is inconsistent with the GRMHD distribution (right [Figure 3.6](#)), which is uniform in ξ_τ . The uniform distribution was the expected result since each simulation is randomly rotated before imaging. The north-south bias is similar to the results found for the elliptical ring in [Section 3.3](#), and the circular crescents in [Figure 2.4](#).

Taking the measured ellipticity and orientation bias together suggests that the imaging algorithms create a preferred ellipticity orientation, and along this direction the ellipticity uncertainty is maximized. This is shown in the bottom panel of [Figure 3.6](#). Here we see that when the ellipticity orientation is aligned in the N-S direction, the recovered ellipticity is very uncertain, and can be quite different from the truth. To add this ellipticity to the results for M 87 we take each recovered ellipticity and orientation from the M 87 topset and add the theoretical uncertainty. This gives that the ellipticity of M 87 is $\tau \in [0.0, 0.3]$. This limit however assumes that M 87 can be approximated by a GRMHD simulation. However, as was shown in [Section 3.3](#), the `eht-imaging` top set can also suppress ellipticity. In the next section we will analyze what happens for the stretched set of GRMHD simulations that include additional ellipticity.

3.4.5 Stretched set results

While the results from the scaled set of GRMHD simulation suggest an upper bound of $\tau \lesssim 0.35$, it does not answer what happens when additional ellipticity, not due to the accretion disk, is in the image. To test this, we will use the GRMHD stretched data set described above. To extract ellipticity and orientation from the stretched set, we again used VIDA’s `CosineRing{1,4}` template with a constant background and the Bh divergence. However, we found that a small subset of the image reconstructions had an additional circular blob present. Due to the second component, VIDA would sometimes report an artificially high ellipticity since the template would try to cover both the central ring and Gaussian blob. To remedy this issue, we added a Gaussian component to the template.

The residuals for the ellipticity position angle are shown in [Figure 3.7](#). Note that we only show the results for VIDA since REx is unable to measure the position angle of the ellipticity. Looking at the bottom panel we see that for intrinsic $\tau \lesssim 0.2$, the position angle residuals are very broad, and can be significantly biased from zero. However, as the intrinsic τ increases we find that the residuals improve. This bias for smaller ellipticity is due to ξ_τ being strongly biased towards 0° . This bias continues as τ increases but its impact is lessened.

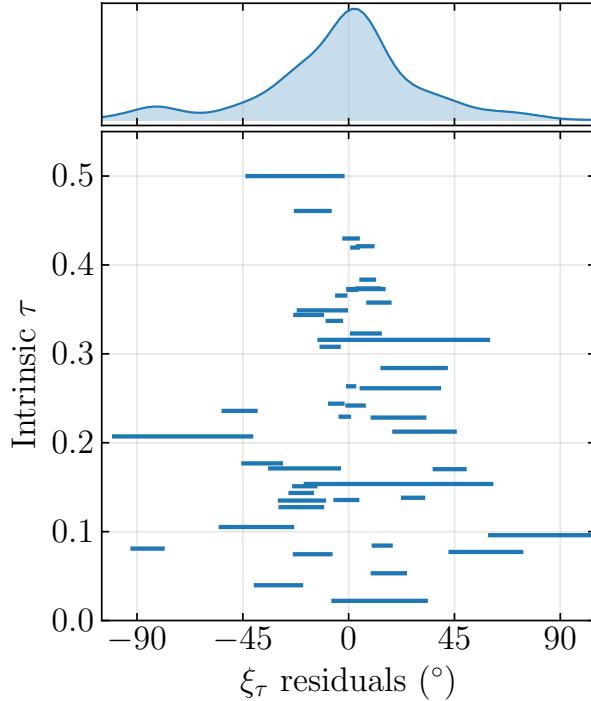


Figure 3.7: Residual distributions of the stretched GRMHD data sets for the ellipticity positions angle ξ_τ . The top panel shows the marginal distribution of the ellipticity residual across the entire stretched GRMHD set. The bottom shows the 95% probability interval of the position angle residuals for each simulation, separated by the intrinsic τ on the y-axis. The intrinsic τ and ξ_τ was found by applying VIDA to the GRMHD snapshots blurred by a Gaussian with a FWHM of $10\mu\text{as}$. The large residuals for small τ are a result of the ξ_τ being heavily biased north-south, similar to the elliptical ring, and scaled GRMHD results.

The left panel of Figure 3.8 shows a map from the recovered τ on the x-axis and the intrinsic τ . The horizontal bars are the 95% confidence regions about the median of the top set images. From this we see that the recovered ellipticity is largely independent of the intrinsic ellipticity when the intrinsic $\tau \lesssim 0.325$. Furthermore, even a GRMHD simulation with intrinsic $\tau = 0.475$, can have a recovered $\tau = 0.1$, which is similar to the value obtained for the M 87 top set.

To quantify the the maximum allowed ellipticity that is consistent with M 87 we first made two cuts on the stretched GRMHD reconstructions. Namely, we remove any reconstructions where $\tau > 0.2$ and $\xi_\tau \notin [-75^\circ, 0^\circ]$. The remaining simulations, therefore, match

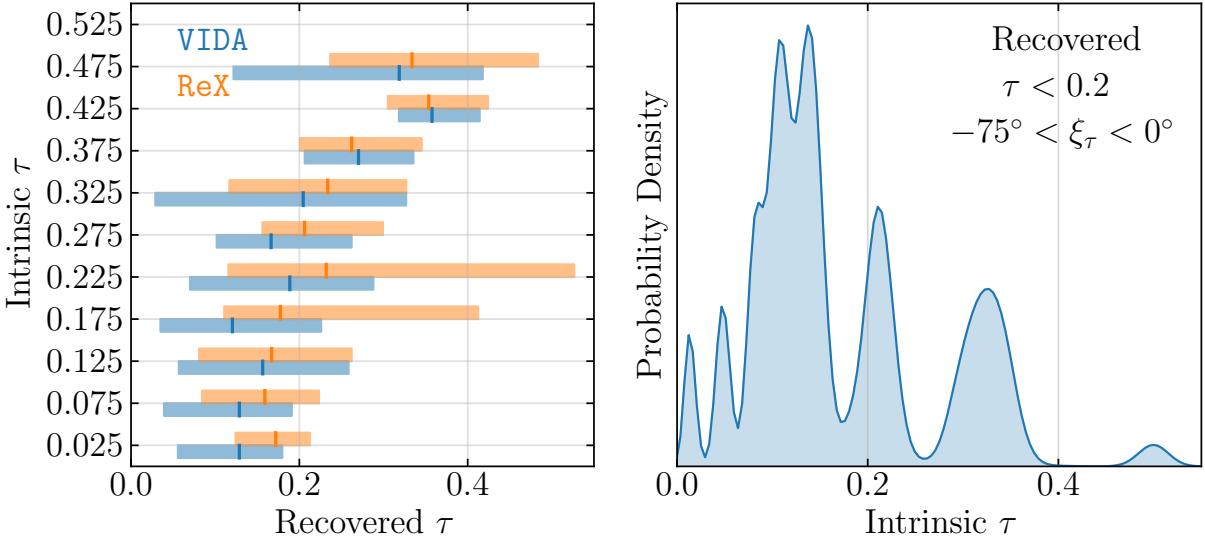


Figure 3.8: Left: Intrinsic ellipticity τ vs recovered ellipticity of the stretched GRMHD images. The intrinsic ellipticity was found by applying VIDA to the GRMHD snapshot images blurred with a $15\mu\text{as}$ Gaussian kernel. The solid line shows the median recovered ellipticity from VIDA (blue) and REx (orange). The filled in bars are the 95% confidence intervals around the median. Right: The resulting intrinsic ellipticity distribution (measured by VIDA), for reconstructions with reconstructed ellipticity and orientation similar to the observed M 87 top set values.

the M 87 top set measurements. The intrinsic ellipticity distribution from the remaining simulations is shown in the right panel of Figure 3.8. From this distribution we see that stretched simulations with ellipticity as high as 0.5 can have ellipticity similar to the observed M 87 results. It is important to note that this is the highest intrinsic ellipticity considered in the stretched set.

3.5 Summary and Conclusions

Measuring the ellipticity of the accretion flow around M 87 is a theoretically interesting property related to the nature of the accretion flow and structure of spacetime. While the results in EHTC VI measured an ellipticity in the image reconstruction of M 87, no attempt was made to interpret or calibrate this result. This chapter has shown that the

top set used for the M 87 images cannot directly measure the on-sky image ellipticity. We demonstrated that even for simple geometric models, the `eht-imaging` top set failed to recover the correct ellipticity in 8/10 test cases.

To account for ellipticity bias, we calibrated the M 87 ellipticity using a set of 550 GRMHD images. Assuming that the ellipticity in the reconstructions is due to accretion turbulence, we found that accounting for the imaging bias the ellipticity of M 87 could be anywhere from $\tau = 0$ to $\tau = 0.3$. However, if there is additional non-accretion induced ellipticity from, for example, some non-GR effect, we found that ellipticity as high as $\tau \approx 0.5$ could have a recovered ellipticity similar to the M 87 results.

The origin of this bias is twofold. First, the EHT coverage in the north-south direction has a large gap around $3 - 5G\lambda$, which is roughly where the `eht-imaging` top set typically tries to place the ellipticity. In this region `eht-imaging` relies on its regularizers to fill in the uv-plane. One of the regularizers used in the `eht-imaging` top set is the *maximum entropy regularizer*:

$$\mathcal{R}_{MEM}(I||P) = - \sum_{ij} I_{ij} \log \left(\frac{I_{ij}}{P_{ij}} \right), \quad (3.2)$$

where P_{ij} is some prior image chosen to approximately match the expected image. For `eht-imaging`, this prior image was a symmetric Gaussian. This prior image will prefer more symmetric images, especially when the ellipticity is in a region of poor coverage. Second, and more fundamentally, the reason for the ellipticity bias is that parameter surveys do not form an image posterior. Instead, they rely on training sets to decide which set of images meet some heuristic threshold and can fail when applied to images not originally considered. To overcome this failure, a statistically rigorous method to assign uncertainty is required. In the next two chapters we will use THEMIS's Bayesian imaging to rigorously measure the ellipticity of M 87.

Chapter 4

Next Generation Samplers for Themis

4.1 Introduction

In the previous chapter, we demonstrated that the imaging techniques used in EHTC IV were unable to recover ellipticity even for simple geometric models. The reason for this is due to image reconstruction uncertainty. The EHT array is very sparse, meaning that many different image morphologies can explain observations. Classical imaging techniques such as RML and CLEAN are unable to measure this uncertainty. However, as explained in Chapter 1, Bayesian methods provide a mathematical framework to assign an uncertainty to image reconstructions.

Bayesian imaging requires sampling from high dimensional multi-modal parameter spaces. Unfortunately, standard MCMC methods such as Gaussian random walks and ensemble methods like emcee (Foreman-Mackey et al., 2013) suffer from the “curse of dimensionality”. While the methods originally implemented in Broderick et al. (2020), were able to sample from distributions with modest dimensionality $\lesssim 25$, their performance degrades in higher dimensions (Huijser et al., 2015). Hamiltonian Monte Carlo (HMC) is a state-of-the-art MCMC method that forms the basis of the popular statistical language Stan (Carpenter et al., 2017). However, implementing these new samplers in THEMIS was difficult since no formal interface existed. Furthermore, THEMIS samplers assumed we used ensemble methods, making it time consuming to implement other methods such as HMC. Motivated by these issues, in this chapter we develop a new sampler framework in THEMIS. This framework defines a common abstract interface for THEMIS samplers. Using

this framework, we developed an interface between Stan and THEMIS, and added a new slice sampler targeted for low dimensional problems.

While HMC gives state-of-the-art performance for local posterior exploration, it struggles to sample the parameter space globally in the presence of multiple modes. To solve this global sampling problem typically requires additional algorithms to move between different modes. One class of algorithms uses prior knowledge of the posterior modes locations and then constructs swaps that preserve detailed balance (Pompe et al., 2020, e.g.). However, these methods require finding the different modes before sampling which can be just as difficult as sampling (Ma et al., 2019). An alternative algorithm is parallel tempering (PT) (Swendsen & Wang, 1986; Hukushima & Nemoto, 1996). THEMIS previously used a PT algorithm from Vousden et al. (2016). While the tempering scheme worked well for some test problems, we often found that it was suboptimal. That is, we would often not efficiently use all the tempering levels. Furthermore, the Vousden et al. (2016) scheme only works for ensemble-based samplers, preventing its use with more general samplers. To overcome these issues, we implemented a new state-of-the-art parallel tempering scheme from Syed et al. (2019).

This chapter will first review THEMIS’ new sampler interface, and detail the two new local samplers implemented. In the next section, we will review parallel tempering and the optimal scheme from Syed et al. (2019). Finally, we will benchmark our new samplers on the Bayesian imaging problem for M 87.

4.2 Abstract Sampler Framework

In (Broderick et al., 2020), the samplers were treated differently than the rest of THEMIS. Namely, while the model and likelihood framework had a single abstract class that defined several primitive functions, THEMIS samplers had to be re-implemented from scratch when a modification was required. This is unfortunate since large parts of the implementations were identical. Moreover, copying over the redundant code from one sampler to another is prone to subtle bugs. To alleviate these issues and simplify the process of adding new samplers, we rewrote the sampler interface.

Samplers are now defined by a single abstract class, called `sampler_MCMC_base.h`. By default, this class handles opening and closing output streams, setting MPI communication, and post-processing like finding the best fit from an MCMC run. On top of this, it defines the interfaces that a user must implement when adding new samplers:

- `set_initial_location`: specifies starting point of the chain.

- `get_chain_state`: gets the current state of the sampler, i.e., the position in parameter space and the current settings of the sampler.
- `set_chain_state`: set the state of the sampler, i.e., the position in parameter space and the current settings of the sampler.
- `write_state`: write the state of the sampler to the file. This function depends on the sampler, e.g., HMC requires the mass matrix and other adaptation parameters.
- `write_checkpoint`: writes the current state to a checkpoint file.
- `read_checkpoint`: reads the checkpoint file and resumes the sampler.
- `run_sampler`: runs the sampler.

The benefit of this approach is that we could write a generic parallel tempering sampler that does not know the particulars of the local sampler. Therefore, adding a new local sampler is decoupled from tempering. This implies any local sampler that satisfies this interface can easily be used with THEMIS' tempering algorithm.

4.3 Local Samplers

4.3.1 Hamiltonian Monte Carlo

THEMIS previously used ensemble-based samplers. More specifically, the affine invariant MCMC method (Goodman & Weare, 2010) and the differential evolution sampler (Das & Suganthan, 2011; Nelson et al., 2014). Unfortunately, ensemble-based samplers can scale poorly to high dimensions (Huijser et al., 2015). Specifically, ensemble samplers tend to have a long burn-in time that can be deceiving with just visual inspection of trace plots. One of the problems with standard Gaussian random walk MH strategies, like ensemble methods, is that these samplers struggle to take large steps in high dimensions. Theoretical arguments (Roberts & Rosenthal, 2001) show that the optimal step size for RW MH goes as $2.38/\sqrt{d}$, meaning that step size actually decreases as the dimension increases. As a result, the sampler tends to slowly diffuse around the parameter space. Hamiltonian Monte Carlo (HMC) however can take larger steps and scales much better in dimension (Beskos et al., 2013). This is accomplished by using deterministic proposals that greatly diminishes the diffusive nature of the sampler (Betancourt, 2017).

The first step in HMC is to introduce N new auxiliary parameters called momenta p , where N is the number of parameters θ . This gives a new posterior $\pi(p, \theta) = \pi(p|\theta)\pi(\theta)$, where $\pi(\theta)$ is the posterior of our original problem. For standard HMC, $\pi(p|q) = \pi(p)$. Furthermore $\pi(p)$ is typically chosen to be a multivariate Gaussian with some constant covariance matrix Σ . The extended posterior of the (p, θ) system is then defined by $\pi(p, \theta) \propto e^{-H(p\theta)}$, where H , the Hamiltonian, is given by

$$\mathcal{H}(p, \theta) = \frac{1}{2}p^T \Sigma^{-1} p + V(\theta), \quad (4.1)$$

where $V = -\log \mathcal{L}(V|\theta)p(\theta)$, is the negative log-joint distribution. HMC proposes a new position (p', θ') in two steps:

1. At the current positions θ , generate a random momentum by drawing it from the distribution $\pi(p)$
2. Propose the new position θ_T, p_T by solving Hamilton's equations

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial \theta}. \quad (4.2)$$

Interestingly, if we could exactly solve Hamilton's equations then this is a valid MCMC scheme, i.e. no acceptance step is needed. This is due to the fact that Hamilton's equation preserve the Hamiltonian and phase space volume through Liouville's theorem. To see that why no MH step is needed we start with the condition that any valid MCMC algorithm must preserve π , i.e.

$$\pi(p', \theta') = \int t(p'\theta'|p, \theta)\pi(p, \theta)d\theta dp, \quad (4.3)$$

where $t(\theta'|\theta)$ is our transition kernel. The HMC transition kernel is

$$t(p', \theta'|p, \theta) = \delta(p' - \phi_T^p(p, \theta))\delta(\theta' - \phi_T^\theta(p, \theta)), \quad (4.4)$$

where $\phi_T^{\theta,p}(p, \theta)$ is the flow from solving Hamilton's equation forward for T units, starting from (p, θ) . Substituting this into (4.3) gives the condition

$$\pi(p, \theta) = \pi(\phi_T^{-1}(p, \theta)) \left| \frac{\partial \phi_T^{-1}(p, \theta)}{\partial(p, \theta)} \right|, \quad (4.5)$$

which is exactly satisfied since H is preserved and the Jacobian term is unity.

Unfortunately, Hamilton's equations cannot be solved analytically. Instead symplectic solvers, such as the Leapfrog method, are used. These methods still preserve phase space volume, implying the Jacobian term in (4.5) is unity. However, the Hamiltonian is no longer preserved. To correct for the change in energy, a MH acceptance step is required to enforce that HMC preserve samples from π :

$$\alpha = \min \left(1, e^{-(H(p', \theta') - H(p, \theta))} \right). \quad (4.6)$$

HMC has several tuning parameters that are essential for good performance. The first two are the step size of the symplectic solver ϵ and the total integration time $T = N_l \epsilon$, where N_l is the number of leapfrog steps. If ϵ and N_l are too small, then HMC devolves to standard random walk MH. Conversely, if N_l and ϵ are too large, the MH acceptance ratio becomes small since ΔH becomes large due to leapfrog integration errors.

Generally, no constant N_l will work for complicated distributions (Betancourt, 2017). Modern HMC samplers instead use dynamic methods to find the optimal number of leapfrog steps at each HMC transition. A popular current method is the no u-turns or NUTS sampler (Hoffman & Gelman, 2011). NUTS works by creating a tree of leapfrog steps both forwards and backwards from the initial position p, θ until the NUTS condition is met:

$$\begin{aligned} p_+^T \cdot (\theta_+ - \theta_-) &< 0 \\ p_-^T \cdot (\theta_- - \theta_+) &< 0, \end{aligned} \quad (4.7)$$

where p_{\pm}, θ_{\pm} are the beginning and ends of the trajectories. Conceptually, this approximately gives the longest possible trajectory before the curve starts to close. For a more detailed review of HMC and NUTS see Betancourt (2017). To find the best ϵ a Nesterov-dual averaging scheme is used (see e.g., Hoffman & Gelman, 2011, for the details).

HMC also benefits from adapting the covariance matrix Σ of $\pi(p)$, which is known as preconditioning. Tuning Σ , N and ϵ together in an efficient manner is challenging. Instead of developing our own method, we used the NUTS sampler from the probabilistic programming language Stan (Carpenter et al., 2017). Stan is written in C++, which allowed us to easily incorporate its sampler into the MCMC framework described in Section 4.2. Currently we have interfaces to Stan's NUTS sampler with a dense¹, and a diagonal momentum covariance matrix². Following the Stan team's suggestion, we mostly use the diagonal covariance matrix. While the dense covariance matrix can model a linear parameter correlation, it is susceptible to numerical issues and can take a substantially longer to

¹sampler_stan_adapt_dense_e_nuts_MCMC

²sampler_stan_adapt_diag_e_nuts_MCMC

tune. Additionally, in our experience, diagonal NUTS and tempering can efficiently handle linear correlations.

4.3.2 Automated Factor Slice Sampler

In addition to Stan’s HMC sampler, we also added an adaptive slice sampler from Tibbits et al. (2014), known as the automated factor slice sampler (AFSS). Slice sampling is similar in spirit to rejection and Gibbs sampling. However, unlike rejection sampling it does not require a reference or bounding distribution to sample from. For a thorough introduction see e.g., Neal (2003).

To describe slice sampling, we first consider a 1-dimensional distribution with an unnormalized probability density $\gamma(x)$. The slice sampling algorithm is:

1. Evaluate γ at the current parameter location x .
2. Generate a uniform random number between in $[0, \gamma(x)]$, which we call h
3. Find the region, $R = \{x \mid \gamma(x) \geq h\}$
4. Uniformly sample from R giving your new sample x' .

The difficulty with slice sampling is finding the region R in step 3. To find R , we use the methods outlined in (Neal, 2003), called the expand and shrink procedure. This consists of two steps:

- **Expand:**

1. Randomly drop a line-segment of length w around the initial point x
2. Evaluate γ at the line-segment endpoint x_l, x_u .
3. If $\gamma(x_l) < h$ and $\gamma(x_u) < h$ proceed to the shrink step
4. Otherwise add another wedge of length w to the endpoint and return to step 2.

- **Shrink:**

1. Given the wedge $[x_l, x_u]$ produced from the expand step, uniformly sample between x_l and x_h , giving the proposed sample c
2. If $\gamma(c) > h$ we are done. Otherwise set the closer endpoint to c , i.e. x_l or x_h , and repeat step one of the shrink procedure.

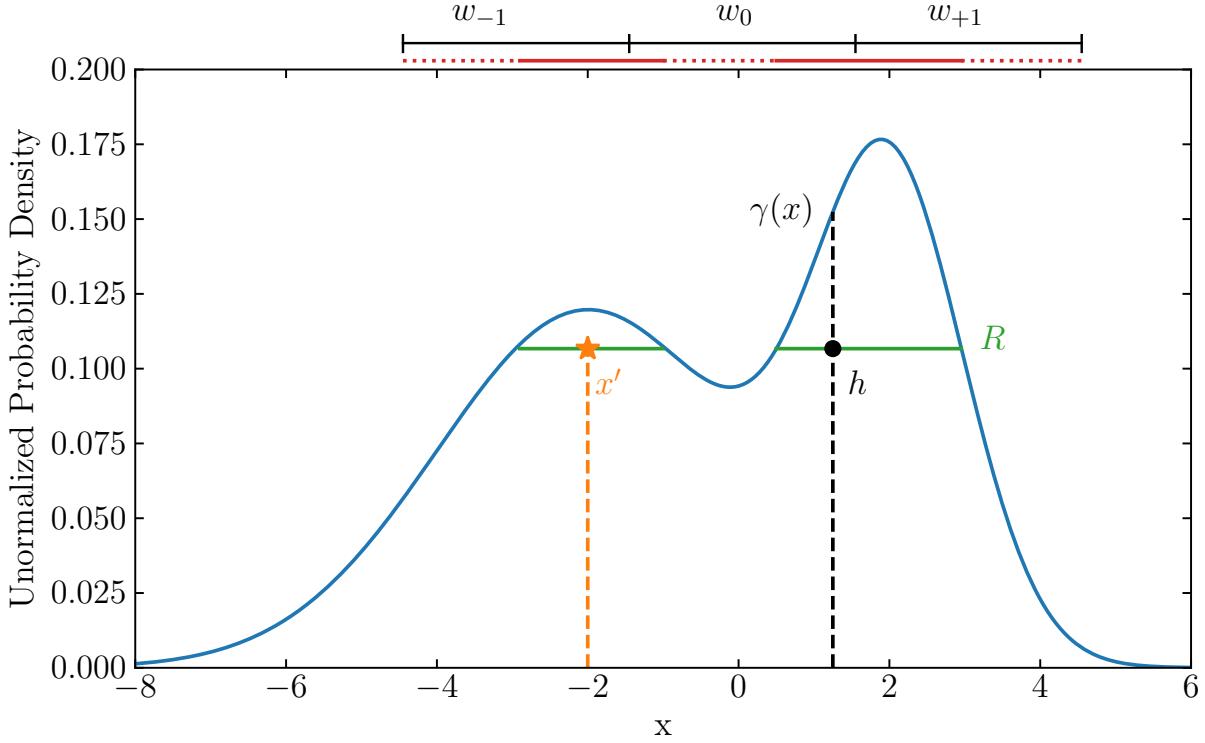


Figure 4.1: Pictorial description of the slice sampling algorithm, using the expand and shrink procedure from [Neal \(2003\)](#). For this step two expansions were required to find the region R , one on each endpoint of the initial wedge.

This procedure is detailed graphically in [Figure 4.1](#). How to choose the initial widths will be detailed below. To generalize this to higher dimensions we iterate through each parameter in the model. This is similar to Gibbs sampling. An added benefit of this approach is that if certain parameters are cheaper to evaluate, we can sample from them more frequently.

Like with HMC, slice sampling benefits greatly from adaptive tuning. For slice sampling, we are concerned with two sets of tuning parameters. The first set is how to choose the initial widths. If the initial widths are poorly chosen, then likelihood evaluations will be wasted on constructing the regions R . Second is preconditioning the slice directions. This preconditioning is similar to the momentum covariance matrix tuning in HMC. For our adaptation strategy we mostly follow [Tibbits et al. \(2014\)](#).

To find the optimal widths, [Tibbits et al. \(2014\)](#) argue that the optimal value initial

width is when the average number of expand N_{expand} and shrink steps N_{shrink} satisfies

$$\frac{\langle N \rangle_{\text{expand}}}{\langle N \rangle_{\text{shrink}} + \langle N \rangle_{\text{expand}}} = \frac{1}{2}. \quad (4.8)$$

To find the initial widths in each direction that satisfy this relation, the sampler is run for some number of steps N_w , where we record the total number of expansions and contractions. After N_w steps the widths are updated according to

$$w' = 2w \left(\frac{\langle N \rangle_{\text{expand}}}{\langle N \rangle_{\text{shrink}} + \langle N \rangle_{\text{expand}}} \right). \quad (4.9)$$

This satisfies a Robbins-Monroe condition on (4.8), and therefore is guaranteed to converge to the optimal widths.

For preconditioning, we first find the Gaussian approximation of the posterior using the current samples to estimate the covariance matrix. Note that we use the online covariance estimator

$$\begin{aligned} \Delta_{n+1} &= \theta_{n+1} - \hat{\theta}_n \\ \hat{\theta}_{n+1} &= \frac{n\hat{\theta}_n + \theta_{n+1}}{n+1} \\ \Sigma_{n+1} &= \frac{n-2}{n-1} \Sigma_n + \frac{1}{n} \Delta_n \Delta_n^T. \end{aligned} \quad (4.10)$$

This prevents us from having to keep the samples in memory. Given this covariance matrix estimator, we find the rotation matrix that diagonalizes it. Given this rotation matrix, we can then orient the slice directions to coincide with the principal axes of the covariance matrix. The benefit of this approach is that after rotation, all linear correlations between the parameters have been removed. Namely, if the posterior is a correlated multivariate Gaussian, after applying the rotation, it is uncorrelated. Note that if there are non-linear correlations, this adaptation strategy may struggle. In those cases, we have found that the NUTS sampler performs better.

Both of these additions to the slice sampler typically require online adaptation. Similar to the Stan adaptation, we break it into two steps. At the beginning of sampling, we update the widths every 10 MCMC steps. We call this the fast adaptation round, and it quickly updates the widths to reasonable values. After the fast adaptation, we update the widths using a doubling time scale. That is, the next update is after 20 steps, and then the next is 40 steps. This reduces the Monte Carlo error in the estimates for (4.8). Finally, after four times the number of parameters, we construct the covariance matrix from the

previous samples. At this point, we then rotate the directions of the slices and restart the fast width adaptations given the new slice directions. This process is then repeated, but with the number of steps until the subsequent covariance matrix adaptation doubled. We repeat this procedure until a user-specified number of adaptation steps.

The reason for doubling the number of adaptation steps and then terminating the adaptation is two-fold. First, constantly adapting an MCMC sampler changes the stationary distribution (Roberts & Rosenthal, 2009; Robert & Casella, 2013), and either the adaptation has to decrease geometrically or terminate during the run to ensure you are sampling from the correct distribution. Second, early in the run, the estimates for the covariance matrix will be noisy. However, not adapting at all during the beginning can also harm performance. Thus, by combining a geometric increase and terminating, we get the benefits of both approaches.

4.4 Global Samplers: Parallel Tempering

In the previous section, we introduced the general sampler class in THEMIS, and two new local samplers. However, neither sampler can easily move between different modes. In this case, a global sampler that explores the entire surface is needed. To solve this, we will use parallel tempering (PT) which was independently developed in both physics (Swendsen & Wang, 1986; Hukushima & Nemoto, 1996) and statistics (Geyer, 1991). PT was used in both of THEMIS’ previous samplers and has been used extensively in astrophysics (e.g., De et al., 2018; Ashton et al., 2019; Biwer et al., 2019).

4.4.1 Review of Parallel Tempering

PT attempts to sample from the difficult posterior distribution $\pi(\theta)$ by introducing a path from a easy to sample or reference distribution π_0 to π using,

$$\pi^{(\beta)}(\theta) = \left(\frac{\pi(\theta)}{\pi_0(\theta)} \right)^\beta \pi_0(\theta) = L_\beta(\theta) \pi_0(\theta), \quad (4.11)$$

where $\beta \in [0, 1]$, and $L_\beta(\theta) = (\pi/\pi_0)^\beta$. Note that sometimes β is written as $1/T$, making a connection to the Boltzmann distribution and thermodynamics. Typically π_0 is taken to be some simple distribution or the prior. The power of parallel tempering is that the distributions with $\beta \sim 0$ are assumed to be much easier to sample since they are “close” to the reference distribution π_0 . The low β distributions then act as a bridge that allow

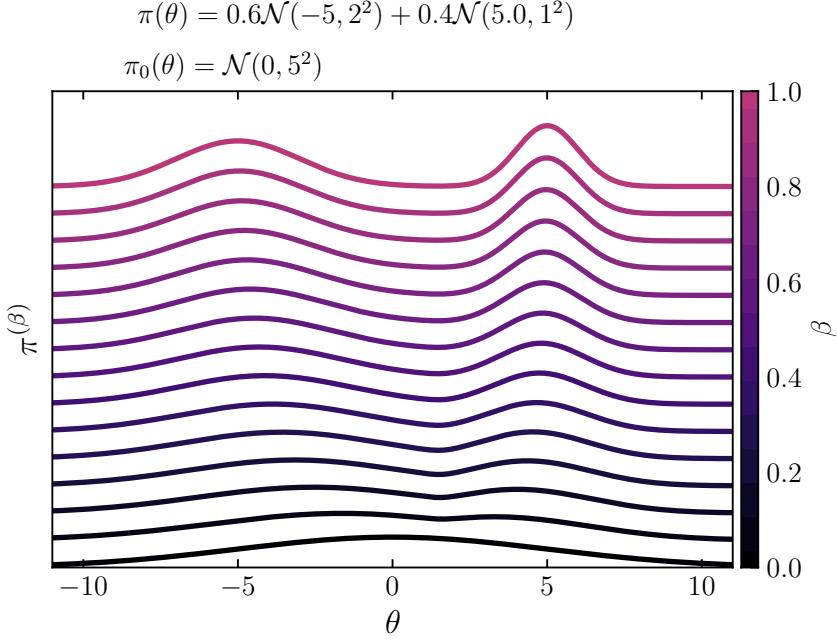


Figure 4.2: Example of how tempering interpolates between the hard to sample distribution π and the reference π_0 , for a Gaussian mixture model.

the sampler to move through low probability regions. This is shown in Figure 4.2 for a Gaussian mixture model.

While in principle one could try to sample the continuous β variables this is quite difficult (Betancourt, 2015). Instead, parallel tempering first partitions β such that $0 = \beta_0 < \beta_1 < \dots < \beta_{N_T-1} < \beta_{N_T} = 1$. This partition, or *ladder*, will be denoted by \mathcal{P} . This gives $N_T + 1$ distributions we need to sample from. Or equivalently we need to sample the tuple $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{N_T})$ whose distribution is

$$\Pi(x) = \prod_{i=0}^{N_T} \pi^{(\beta_i)}(\theta_i) = \prod_{i=0}^{N_T} L_{\beta_i}(\theta) \pi_0(\theta). \quad (4.12)$$

To sample from Π requires two independent phases called *exploration* and *communication*. During the exploration phase, we independently apply a local MCMC kernel for each β_i . This process is trivially parallelizable, giving PT its name. After the exploration phase, we need to communicate between different chains. This communication is achieved

by swapping between different β_i :

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}_{i,j} = (\theta_0, \theta_1, \dots, \theta_j, \theta_{i+1}, \dots, \theta_i, \theta_{j+1}, \dots, \theta_N). \quad (4.13)$$

These swap are accepted with a probability

$$\alpha^{(i,j)} = \min \left\{ 1, \frac{\Pi(\boldsymbol{\theta}_{i,j})}{\Pi(\boldsymbol{\theta})} \right\} = \min \left\{ 1, \left(\frac{L_{\beta_i}(\theta_i)}{L_{\beta_j}(\theta_j)} \right) \right\}. \quad (4.14)$$

Which indices (i, j) are swapped are left as a design decision when implementing parallel tempering. In [Section 4.4.2](#), we describe the swapping kernel we use in THEMIS.

4.4.2 Optimizing Parallel Tempering

From [Figure 4.2](#), we see that tempering allows one to move to different modes by traversing from $\beta = 1$ down to $\beta = 0$, and back up to $\beta = 1$, which is commonly called a round trip. Therefore, to optimize parallel tempering, we seek to minimize the number of communication steps n that it takes for a round trip to complete. In order to minimize the round trip rate, τ , we consider two design choices:

1. Efficient communication or swapping strategy
2. Good partition or ladder, \mathcal{P} , of β

To optimize the two points we follow [Syed et al. \(2019\)](#) which, under some mild assumptions about π , is able to derive the optimal discretization of a certain class of non-reversible parallel tempering schemes applicable to distributed computing environments.

Communication Strategy

THEMIS is typically run on large HPC environments typically composed of thousands of independent CPUs. Therefore, to reduce communication overhead, we will restrict ourselves to swapping schemes that do not have clashes. Namely, swapping only requires communication between two levels. With this restriction, there are two swapping kernels we can consider:

$$K^{\text{even}} = \prod_{i \in \text{even}} K^{(i,i+1)}, \quad K^{\text{odd}} = \prod_{i \in \text{odd}} K^{(i,i+1)}. \quad (4.15)$$

Therefore, to swap the entire ladder, we then need to compose these swaps. One method is to randomly choose K^{even} or K^{odd} at each swap. This strategy preserves detailed balance and is called *stochastic even-odd swap* (SEO). However, as is shown in Syed et al. (2019), SEO suffers from a similar diffusion problem as in random-walk MH. To overcome this limitation, THEMIS uses the *Deterministic even odd* (DEO) swapping scheme:

$$K_n^{\text{DEO}} = \begin{cases} K^{\text{even}} & \text{if } n \text{ is even,} \\ K^{\text{odd}} & \text{if } n \text{ is odd,} \end{cases} \quad (4.16)$$

where n is the current swap number. Notice that this kernel is not reversible. Namely, if the communication step i uses an even swapping kernel, then for the next swap, there is no way for the sampler return to its previous state. To see the benefit of this approach, we first define several necessary quantities that assess the performance of parallel tempering.

Let the $r(\beta_i, \beta_j)$ be the average rejection probability between x_i, x_j , i.e.

$$r(\beta_i, \beta_j) = 1 - \langle \alpha^{(i,j)} \rangle. \quad (4.17)$$

Then Syed et al. (2019) showed that for finite N_T ³, the round trip rate goes as

$$\tau = \frac{1}{2 + 2E(\mathcal{P})}, \quad E(\mathcal{P}) = \sum_{i=1}^{N_T} \frac{r^{(i,i-1)}}{1 - r^{(i,i-1)}}. \quad (4.18)$$

This is a direct consequence of using the DEO swapping scheme. If we had used the reversible SEO tempering scheme, then τ goes as $(2N_T + 2E(\mathcal{P}))^{-1}$ (Syed et al., 2019). Therefore, for large enough N_T , the performance of SEO decreases, due to its diffusive properties. Equation 4.18 also implies that as long as the number of swaps is much larger than the number of tempering levels, the DEO's round-trip rate monotonically improves with more tempering levels. That is, unlike SEO we never harm the performance of DEO by increasing the number of tempering levels, as long as the number of swaps is much larger than the number of tempering levels. However, while this motivates the use of DEO, we still need to find the optimal partition \mathcal{P} .

Optimal Ladder

Equation 4.18 also provides an avenue to optimize the ladder choice \mathcal{P} . Namely, by minimizing $E(\mathcal{P})$ we can increase the round trip rate. Intuitively, we can see that $E(\mathcal{P})$ will

³This result does assume that the number of swaps or communication steps is much greater than the number of tempering levels.

decrease as $N \rightarrow \infty$, since $\alpha^{(i,j)} \rightarrow 1$. Therefore, to optimize $E(\mathcal{P})$ we first turn to the limit where the number of tempering levels is large or equivalently as the spacing goes to zero.

In this limit, there are two important quantities $\lambda(\beta)$ and Λ :

$$\lambda(\beta) = \frac{d}{d\beta'} r(\beta, \beta') \Big|_{\beta'=\beta}, \quad \Lambda(\beta) = \int_0^\beta \lambda(\beta') d\beta'. \quad (4.19)$$

λ and Λ are called the local and global communication barriers respectively.

In the high N limit Syed et al. (2019) proved that the round trip rate for DEO approaches

$$\tau^{\text{opt}} = \frac{1}{2 + 2\Lambda(1)}, \quad (4.20)$$

and that it approaches it monotonically from below, i.e. τ_{opt} is optimal. $\Lambda(1)$ then measures the total communication bottle neck of the problem, and measures how similar the reference distribution π_0 is to the posterior π .

Following Syed et al. (2019), we can then optimize our ladder partition by finding a \mathcal{P} such that $\Lambda(\beta) \approx E(\mathcal{P})$. For this we will first need the following result from (Syed et al., 2019):

$$\Lambda(\beta_j) \approx \sum_{i=1}^j r(\beta_{i-1}, \beta_i), \quad (4.21)$$

for any \mathcal{P} . Finding the optimal ladder then amounts to minimizing $E(\mathcal{P})$ subject to the constraint that $\sum_{i=1}^N r(\beta_{i-1}, \beta_i) = \Lambda(1)$. This problem can easily be solved using Lagrange multipliers for $r(\beta_{i-1}, \beta_i)$. The resulting minimum gives that the rejection rates between tempering levels should be constant.

Since the optimal ladder had constant rejection rates, we have that $\Lambda(\beta_k)$ satisfies

$$\Lambda(\beta_k) \approx \frac{k}{N} \Lambda(1). \quad (4.22)$$

Since $\Lambda(\beta)$ is a monotonically increasing function, we can then solve for each β_k using a root finding algorithm, such as the bisection method. Furthermore, keeping track of $\Lambda(\beta_k)$, just involves storing a single number during sampling, adding negligible computational overhead.

4.4.3 Implementation in Themis

To implement DEO in THEMIS, we used created a subclass of the `sampler_MCMC_base` class called `sampler_deo_tempering_MCMC.h`. To specify the exploration kernel, we use C++ templates to specify the exploration kernel. This allows the user to easily specify different exploration kernels with minimal changes to their THEMIS scripts. Our implementation mostly follows the suggestions made in (Syed et al., 2019). However, we do not swap β between MPI communicators but rather, the parameters θ . This change is mathematically equivalent but makes designing efficient local exploration kernels easier. Namely, we can keep the adaptation schemes from Section 4.3. Additionally, while any reference distributions can be used, THEMIS defaults to set $\pi_0(x)$ to be a uniform hypercube with boundaries specified by the user. In the future other reference distributions may be chosen.

To adapt the ladder we follow (Algorithm 2. Syed et al., 2019), which breaks the a run into tempering *rounds*. The number of swaps characterizes each round and is increased geometrically between rounds⁴.

The output of a THEMIS run is several ASCII text files. In addition, we provide several diagnostic plots to summarize the performance of the ladder. These diagnostics are included with the THEMIS post-processing python package THEMISPy and provides several sampler performance diagnostics.

4.4.4 Comparison to Themis' Previous PT Sampler

THEMIS' tempering scheme differs from the old scheme in two ways. First, the old tempering scheme was restricted to only work with ensemble-based methods. This restriction was a significant motivation to implementing a new tempering scheme. Secondly, the old samplers used a different tuning strategy, based on continuous diminishing adaptation from (Vousden et al., 2016). In practice, we often found that the ladder would not adapt fast enough, and by the time the adaptation had essentially frozen, the ladder would be poorly adapted. Additionally, we often found that the ladder would never extend to low enough β , Namely, we often had instances where most tempering levels would concentrate around $\beta = 1$, essentially removing any effects of tempering. To fix this problem, we adapted the (Vousden et al., 2016) scheme to drive all acceptance rates to 0.5. While this prevented the $\beta = 1$ concentration, we found that instead, large parts of the ladder would peel off to $\beta = 0$. This behavior was because we had effectively over-saturated the ladder, wasting computational resources. With the new Syed et al. (2019) tempering scheme, all these

⁴THEMIS allows you to specify the geometric factor, but in most instances, we set it to 2.

issues have been solved. Additionally, we find that the new adaptation scheme is quite efficient, only requiring 5-6 rounds to converge to constant rejection rates.

4.5 Assessing MCMC Convergence

As part of our new sampler interface, we added a number of MCMC convergence diagnostics to the THEMIS python post-processing pipeline THEMISPy⁵. To assess the local sampling performance, we now included analyses of the integrated autocorrelation time (IACT) of the MCMC chain, the related effective sample size, and the improved split- \hat{R} statistic from (Vehtari et al., 2021).

IACT, is a key quantity of interest in assessing the performance of the sampler. Namely, it is related to the Monte Carlo error that arises from using the chain to estimate expectations of the function f :

$$\mathbb{E}_\pi [f] = \mu_f \approx \frac{1}{N} \sum_t f(x_t). \quad (4.23)$$

From the central limit theorem, the Monte Carlo standard error (MCSE) goes as

$$\text{MCSE}_f = \frac{\sigma_f}{\sqrt{N}}, \quad (4.24)$$

where σ_f is the variance of f , i.e.

$$\sigma_f^2 = \frac{1}{N} \sum_t (f(x_t) - \mu_f)^2. \quad (4.25)$$

However, since samples from an MCMC chain are correlated, (4.24) is not valid. Instead it is replaced⁶ by

$$\text{MCSE}_f = \frac{\sigma_f}{\sqrt{N/\ell_f}}, \quad (4.26)$$

where ℓ_f is defined as

$$\ell_f = 1 + \frac{2}{\sigma_f^2} \sum_{i=1}^N c_f(\tau), \quad c_f(t) = \frac{1}{N-t} \sum_{n=1}^{N-t} (f_n - \mu_f)(f_{n+t} - \mu_f). \quad (4.27)$$

⁵<https://github.com/aeb/ThemisPy>

⁶This formula only holds for MCMC chains that are geometrically ergodic. This is a rather stringent condition that may not hold generically (Robert & Casella, 2013).

The quantity N/ℓ is known as the effective sample size (ESS). Therefore, if ℓ is large, the ESS is small, and the Monte Carlo estimate will be poor. To estimate ℓ we currently provide two independent calculations from [Geyer \(1992\)](#) and [Sokal \(1996\)](#).

The other primary convergence diagnostic we consider is split- \hat{R} . Split- \hat{R} is mainly a convergence diagnostic, providing a guide about whether the chain is approximately stationary or still evolving. It was first proposed in [Gelman & Rubin \(1992\)](#) and uses splits chains to measure between chain variance compared to within chain variance. If both variances are similar, then the chains are considered to be “well-mixed”. More recently, the original split- \hat{R} was improved by using rank-normalized versions of the chain (see [Vehtari et al., 2021](#), for details). Additionally, this version is still defined even if the posterior does not have a first or second moment. This statistic, is the version we use in THEMISPy.

To assess tempering performance, we have added for visualizations in THEMISPy. An example figure is shown in [Figure 4.3](#). For a well-tuned ladder both the left and right panels of [Figure 4.3](#) should be comprised of horizontal lines. The right panel shows the estimate for $\lambda(\beta)$ (4.19), and reports the optimal round-trip rate $\hat{\tau}_{\text{opt}}$ (continuous ladder) and estimated rate for the latest round $\hat{\tau}_{\text{est}}$ in the top right. $\lambda(\beta)$ measures the local communication bottleneck of the ladder. A high λ means that it is more difficult to swap in that region. As a result, the our adaptation scheme will concentrate the partition in that region to smooth it out. In practice, we often find that λ increases as $\beta \rightarrow 0$. This signifies that THEMIS’ current default reference distribution, a uniform box, is the main bottleneck for tempering.

4.6 Bayesian Imaging Validation

Bayesian imaging was one of the main motivations for adding new samplers to THEMIS. While the ensemble sampler with the were used in [Broderick et al. \(2020\)](#), the run-time was significant and required considerable computational resources. Therefore, in this section, we will use the new samplers to explore the Bayesian image posterior of M 87 and compare their efficiency to the ensemble samplers. For this, we will only consider imaging one day, April 11. The results from imaging all days will be presented in the next chapter.

For our image model, we will consider an image with a 5×5 raster, using the bicubic interpolation kernel to make a continuous image (for a thorough description of the model, see [Section 5.2](#)). For this test we will consider four samplers:

1. Differential evolution sampler using THEMIS old adaptation scheme based on the Vosudsen tempering scheme

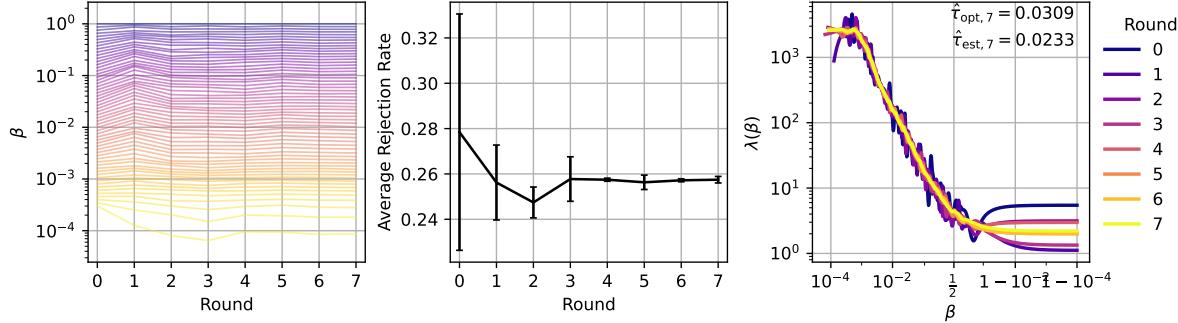


Figure 4.3: An example DEO tempering diagnostic plot taken from the validation Bayesian imaging run in Section 4.6. The left plot shows the evolution of the tempering β parameters over different adaptation rounds. The middle panel shows the average rejection rate and the standard deviation of the rejection rates over each round. The right panel shows the local communication barrier as a function of β across different rounds. In the top right corner is the optimal round-trip rate τ_{opt} , and the estimated round trip rate for using the latest ladder τ_{est} . This plot was produced using THEMISPY.

2. Differential evolution sampler using the DEO tempering scheme
3. Automated factor slice sampler using the DEO tempering scheme
4. NUTS sampler with a diagonal mass matrix and using the DEO tempering scheme

We will use 60 tempering levels for each sampler, where the first round computes 20 swaps, and the number of swaps for each round increases by a factor of 2. The differential evolution and AFSS sampler will use 25 steps per swap, while NUTS will only use ten because it typically takes larger steps than either AFSS or the differential evolution samplers. To ensure computational consistency between methods, we will ran each sampler for 48 hours using 60 cores. While the differential evolution sampler could be further parallelized for each local MCMC move, this is unfair since we could also run additional copies of AFSS and NUTS runs and combine them. All runs are initialized from the same region in parameter space. To compare the results to the truth, we will compare the results to the long runs (over 10 days) used in the next chapter.

The results of the benchmark are shown in Figure 4.4. We found that both the AFSS and NUTS samplers quickly approached the best mode of the posterior in the first 2% of the run. On the other hand, both differential evolution samplers were still approaching

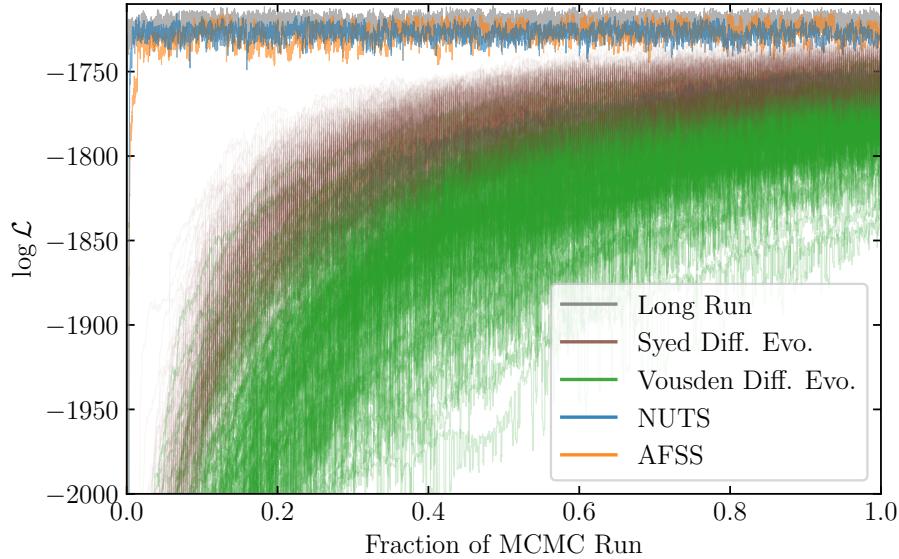


Figure 4.4: Imaging benchmark for the three samplers: AFSS (orange), NUTS (blue), and the ensemble differential evolution sampler using the [\(Syed et al., 2019\)](#) PT scheme (burgundy) and [\(Vousden et al. \(2016\)\)](#) PT scheme (green). We also include the chain from the long-time runs in the next section. All three samplers used the same tempering scheme and were initialized at the same region of parameter space. While the AFSS and NUTS samplers quickly moved to the high probability region, the differential evolution sampler struggled, never exploring the highest likelihood regions.

the peak even after extending the run for an additional 24 hours. This is emblematic of ensemble samplers in higher dimensions. Namely, they tend to have extremely long burn-in times ([Huijser et al., 2015](#)). Comparing the results of the [Syed et al. \(2019\)](#) and [Vousden et al. \(2016\)](#) tempering schemes (grey and purple), we see that the [Syed et al. \(2019\)](#) scheme is moving the peak faster. More quantitatively, the mean log-likelihood values are -1755 v.s. -1796 , and the maximum values are -1728 v.s. -1749 for the [Syed et al. \(2019\)](#) and [Vousden et al. \(2016\)](#) respectively. Given the superior performance of the [Syed et al. \(2019\)](#) tempering scheme, we will not consider the [Vousden et al. \(2016\)](#) scheme in the comparisons below.

To compare the performance of the samplers, we will attempt to find the mean image. However, we found that the resulting image structure was quite multimodal. Therefore, rather than averaging all the modes together, we first used a k-mean clustering algorithm to assign cluster labels. Once we identified the clusters, we then averaged over each image

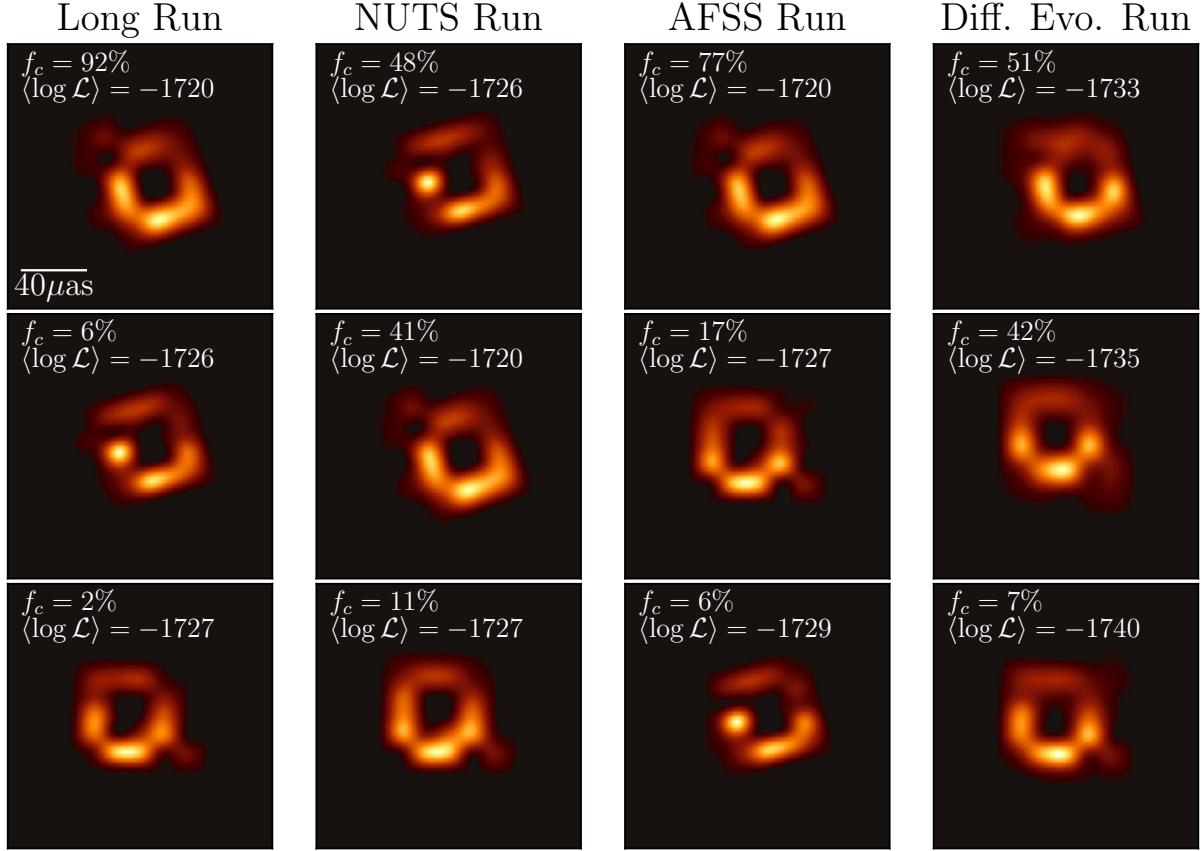


Figure 4.5: Comparison of the performance of the NUTS (second column), AFSS (third column), and differential evolution (fourth column) relative to the “ground-truth” found from a very long run. Given that the posterior is multi modal we took the chain of each run, and identified each image cluster using a k-means algorithm. Overall the NUTS and AFSS samplers are able to find the same three dominant image modes when compared to the long-run. Furthermore, the NUTS sampler has already found the correct average likelihood for the three modes, demonstrating its superior mixing properties. The differential evolution sampler has still not found the three best image structures.

in the cluster. The results of this procedure for the NUTS, AFSS, and differential evolution samplers compared to a long run, are shown in Figure 4.5. Overall we find that NUTS and AFSS were able to identify the three most prominent image structures from the long run. However, the probabilistic mass of each cluster is still evolving to the correct value.

Furthermore, the NUTS sampler has already converged to the correct average likelihood in each cluster. On the other hand, the differential evolution sampler still has not identified the three most dominant clusters, and the average log-likelihood in each cluster is lower by a factor of 10.

This experiment suggests that for Bayesian imaging, the default sampler of choice should be the NUTS sampler, followed by the AFSS. We also found that while the ensemble sampler produces many samples at each “step”, the samples are highly correlated and take a long time to diffuse to the high probability regions of parameter space even with the aid of tempering. This burn-in could be reduced with an optimization algorithm and then starting the sampler from the best fit. However, we found that finding the maximum a posteriori value of the Bayesian imaging problem to be difficult due to its highly multimodal nature.

4.7 Future Improvements

As described in [Section 4.4.2](#) there are three ingredients in parallel tempering:

1. Ladder
2. Communication
3. Reference distribution.

In this work, we have focused on optimizing the first and second points. However, as [Figure 4.2](#) demonstrates, our current communication bottleneck arises from the choice of reference distribution. Namely, as β decreases we see that $\lambda(\beta)$ increases greatly. One way to improve this would be to simultaneously optimize our choice of reference distribution during the tuning phase. For instance, we could use *variational inference* Blei et al. (2017), with some simple variational family to approximate the posterior. Variational inference itself has no guarantees for the asymptotic convergence of expectations of the posterior, e.g., the mean and variance will be biased (sometimes greatly) from the truth. However, we can use this variational approximation of the posterior as the reference distribution and then use parallel tempering exactly sample from the posterior.

The other way to improve parallel tempering is to choose better “paths” between the reference posterior distribution. Parallel tempering takes the linear interpolation path between the log-posterior and log-reference:

$$\log \pi = \beta \log \pi(\theta|V) + (1 - \beta) \log \pi_0(\theta|V). \quad (4.28)$$

By using a finite number of tempering levels we are discretizing this path. The communication strategy detailed in [Section 4.4.2](#) then amounts to finding the optimal discretization given the number of discrete points. In [Syed et al. \(2021\)](#), they demonstrated that the linear path is not optimal even in simple toy models. Instead of picking a linear path we could for instance pick a set of curves $\eta_1(\beta)$ and $\eta_0(\beta)$ such that

$$\log \pi_\beta(\theta|V) = \eta_1(\beta) \log \pi_1(\theta|V) + \eta_0(\beta) \log \pi_0(\theta|V), \quad (4.29)$$

where we require $\eta_1(1) = 1$, $\eta_1(0) = 0$, $\eta_0(1) = 0$, and $\eta_0(0) = 0$.

Another improvement would be to improve the adaptation of the local exploration kernel. While we independently adapt the covariance matrix for each tempering level, this adaptation is not ideal. For instance, if each mode has a different local covariance, the adaptation will struggle to explore each mode efficiently. One way to improve this would be to take inspiration from some of the clustering analysis used in nested sampling ([Feroz et al., 2009](#)). Namely, during the adaptation phases of each tempering round, we could use an online clustering algorithm, such as k-means, to find and track the different clusters. Once assigning the samples to each cluster, we could compute the local covariance matrix of each cluster. We could then switch local covariance matrices depending on the current cluster. This adaptation is similar in spirit to the strategy detailed in ([Pompe et al., 2020](#)), except that we will use parallel tempering to identify the individual clusters.

4.8 Conclusion

In the previous chapter, we demonstrated the need for Bayesian imaging to solve the EHT imaging reconstruction ellipticity problem. However, Bayesian imaging is a complex problem given its high-dimensionality, and potentially multi-modal structure. In this chapter, we have developed a new sampler interface for `THEMIS` and implemented the NUTS and AFSS samplers. Our validation exercise found that this greatly increased our sampling efficiency for Bayesian imaging. Additionally, we built a new optimal tempering scheme that can be used with any local sampler in `THEMIS`.

These developments have already been applied to several analyses of EHT data. For instance, the AFSS was used to measure the non-Gaussian structure of the Sgr A* at 86 GHz, believed to arise from the interstellar scattering kernel ([Issaoun et al., 2021](#)). The NUTS sampler and DEO tempering scheme was also a driving factor in the first polarized images of M 87 ([Event Horizon Telescope Collaboration et al., 2021a](#)). This required sampling models with ~ 350 parameters, which was not possible before the sampler upgrades described in this chapter.

The next chapter will apply this new sampler interface to the M 87 image ellipticity problem. This will allow us to place fully Bayesian ellipticity constraints on M 87, overcoming the challenges detailed in the previous chapter.

Chapter 5

A Bayesian Image Estimate of M 87 ellipticity

5.1 Introduction

While historical imaging methods have been very successful, the sparse coverage of the EHT makes it necessary to understand uncertainty. As a result, there has been a large number of different Bayesian imaging tools developed in the past few years (Broderick et al., 2020; Broderick et al., 2020; Sun & Bouman, 2020; Arras et al., 2021; Pesce, 2021). Each of these tools makes different assumptions about the image structure, and/or, local posterior structure. Sun & Bouman (2020); Arras et al. (2021), both use variational inference to approximate the posterior, but these methods have no guarantee that the expectations from the posterior approximation are reliable (Yao et al., 2018). Broderick et al. (2020) and Pesce (2021) are similar in many regards, in that they use very similar image models and NUTS for local posterior exploration. However, Pesce (2021) does not have parallel tempering implemented, meaning it won't be able to globally explore the parameter space. Therefore, the methods from Broderick et al. (2020) and Chapter 4 should give the most robust estimates.

In Chapter 3, we demonstrated that the traditional imaging tools used for M 87 were unable to quantify the ellipticity of the on-sky image. Namely, we found that the imaging top set failed to recover the correct ellipticity even for simple geometric models. Given this bias, the distribution of measured image reconstruction ellipticity did not represent the true ellipticity of the on-sky image. To overcome this bias, we calibrated the image uncertainty from the top set by using it to image 550 GRMHD simulations. However, this calibration

assumes that the accretion disk in M 87 is similar to the ones in GRMHD simulations. Furthermore, when attempting to place an upper bound in ellipticity we found that images with intrinsic ellipticity as high as $\tau = 0.5$ could have a top set recovered ellipticity similar to the measured M 87 values.

To solve these issues, this chapter will use the sampling tools presented in [Chapter 4](#) to image the M 87 2017 EHT observations using Bayesian techniques. The product of this analysis will be an image posterior. By combining the resulting image posterior with the image feature extraction tool VIDA, we will construct ellipticity posteriors. To demonstrate the validity of this process we will analyze the elliptical ring simulated data from [Section 3.3](#). Unlike the `eht-imaging` top set, we find that Bayesian imaging is able to recover the correct ellipticity structure regardless of the orientation.

Given this result, the ellipticity posterior produced by Bayesian imaging does not have to be calibrated. Therefore, we can directly compare the measured Bayesian ellipticity of M 87 to the expected ellipticity from GRMHD simulations. As a result, we will be able to independently check whether M 87 is consistent with the theoretical expectation from general relativistic fluid dynamics.

The layout of this chapter is as follows: [Section 5.2](#) will review the Bayesian imaging model from [Broderick et al. \(2020\)](#), and the VIDA templates used to extract the ellipticity posterior. [Section 5.3](#) will review the data product we use in this chapter. [Section 5.4](#) will apply Bayesian imaging to the elliptical ring simulated data. Finally in [Section 5.5](#), we measure the ellipticity of M 87 using Bayesian imaging and VIDA.

5.2 Bayesian Imaging and Feature Extraction

5.2.1 Image Domain Model

In [Broderick et al. \(2020\)](#), a new Bayesian imaging (BI) technique was developed that allows for a robust quantification of image uncertainty. The image model starts by assuming that the on-sky image can be described by a set of rectangular control/raster values c_{ij} . This raster is positioned at the locations $(x_i, y_i) = (i\text{FOV}_x/(n_x - 1), j\text{FOV}_y/(n_y - 1))$. The on-sky intensity is given by

$$I(x, y) = \sum_{ij} c_{ij} \kappa(x - x_i) \kappa(y - y_j), \quad (5.1)$$

where κ is an image response function that converts the discrete set raster to a continuous image. The Fourier transform of this image is

$$\tilde{I}(u, v) = \sum_{ij} e^{2\pi i(ux_i + vy_j)} c_{ij} \tilde{\kappa}(u) \tilde{\kappa}(v), \quad (5.2)$$

where $\tilde{\kappa}$ is the Fourier transform of the image response function κ . The choice of κ is an arbitrary choice, but for efficiency should have an analytic Fourier transform. We will follow Broderick et al. (2020), and use a bicubic kernel:

$$\kappa(x) = \begin{cases} 0 & |x| \geq 2 \\ b(|x|^3 - 5|x|^2 + 8|x| - 4) & 1 \leq |x| < 2 \\ (b+2)|x|^2 - (b+3)|x|^2 + 1 & |x| < 1 \end{cases} \quad (5.3)$$

which has Fourier transform

$$\begin{aligned} \tilde{\kappa}(k) = & -\frac{4}{k^3} \sin(k) (2b \cos(k) + (4b+3)) \\ & + \frac{12}{k^4} [b - b \cos(2k) + 2 - 2 \cos(k)]. \end{aligned} \quad (5.4)$$

For a derivation of the Fourier transform, see Appendix A of Broderick et al. (2020). The constant b is a free parameter of the interpolation kernel and is typically set to $b = -0.5$ ¹. This model has been implemented in THEMIS as `model_image_adaptive_spline_raster`.

In Broderick et al. (2020), the grid itself was specified a priori, and then the preferred field of view was found by performing a parameter survey with the field of view and number of raster pixels as free variables. In this chapter, we instead allow the raster size and orientation ξ to vary during sampling. This lowers the number of model hyperparameters to just the number of pixels in the two principle directions.

Furthermore, Bayesian imaging, unlike RML methods, provides a mathematical framework to choose the optimal hyperparameters. To choose the optimal number of raster points, we let the number of pixels to be a model parameter. Therefore, we can form a hierarchical model:

$$p(N_x, N_y, \theta | V) = \frac{p(V | N_x, N_y, \theta) p(\theta | N_x, N_y) p(N_x, N_y)}{p(V)}, \quad (5.5)$$

¹This kernel isn't positive definite so it is possible for the image to have negative intensity. In practice we find that this effect is minimal.

where $\theta = (c_{ij}, \text{FOV}_{x,y}, \xi)$, is the combined raster parameters. Marginalizing over θ gives

$$p(N_x, N_y | V) = \frac{p(V | N_x, N_y)p(N_x, N_y)}{p(V)}. \quad (5.6)$$

Setting the prior $p(N_x, N_y)$ to be uniform over the considered pixel numbers, we see that the model with the preferred model corresponds to model with the highest evidence, $p(V | N_x, N_y)$. Below we will consider a small survey of different pixels resolutions and use the evidence, or rather an approximation of it, to find the optimal N_x, N_y . As a side effect, this survey will also allow us to test whether the Bayesian imaging results are robust to different pixel resolutions.

For our model priors we use a log-uniform distribution between $e^{35} \text{Jy/str}$ and $e^{50} \text{Jy/str}$ on all the raster values, c_{ij} . The raster field of view $\text{FOV}_x, \text{FOV}_y$ were set to be independent with uniform priors over the interval $[15 \mu\text{as}, 120 \mu\text{as}]$ and a uniform prior on the raster orientation. For the gains, we used Gaussian priors on the amplitudes with unit mean and standard deviation 0.1 on all stations except LMT, which used a prior amplitude of 1.0 to model the large pointing offsets during the 2017 observations ([Event Horizon Telescope Collaboration et al., 2019c](#)). The gain phase priors were set to a zero-mean Gaussian with a variance of 10^8 radians.

5.2.2 Image Feature Extraction

The output of Bayesian imaging is a posterior over the image. To move from the image samples to feature posteriors we will use VIDA. We do not include results from REx since, as was demonstrated in Chapters 2 and 3, it gives similar results to VIDA. Furthermore, REx is unable to recover ellipticity orientation which is key in assessing whether the reconstruction is reliable.

For VIDA, we will use the same templates and divergence as in [Chapter 3](#). Namely for the elliptical ring reconstructions we will use the `CosineRing{0,1}` template. For the M 87 reconstructions we will use the `CosineRing{1,4}` template. In all cases we use the Bhattacharyya divergence [\(2.8\)](#). We will mainly be interested in the ellipticity, τ , of the ring and its orientation, ξ_τ , measured east of north.

5.3 Data

The EHT is a very-long-baseline interferometer. By the van Cittert–Zernike theorem ([Thompson et al., 2017](#)) a perfect interferometer measures the Fourier transform of the

on-sky image:

$$\tilde{I}(u, v) = \int I(\alpha, \beta) e^{2\pi i(u\alpha+v\beta)} d\alpha d\beta, \quad (5.7)$$

where $I(\alpha, \beta)$ is the on-sky image intensity. However, in reality the measurements are corrupted by telescope and scan specific complex gain terms $g_i = \gamma_i e^{i\theta_i}$ giving

$$\tilde{V}_{ij} = g_i g_j^* \tilde{I}_{ij}, \quad (5.8)$$

where the \tilde{V}_{ij} are known as the complex visibilities. This is the primary data product we will consider in this chapter. The benefit of using complex visibilities is that the observation likelihood of visibilities is just a complex Gaussian (Thompson et al., 2017)

$$\mathcal{L}(V|\theta) = \prod_i (2\pi\sigma_i^2)^{-1} \exp\left(-\frac{|V_i - g_{1,i}g_{2,i}^* \hat{V}_\theta(u_i, v_i)|^2}{2\sigma_i^2}\right), \quad (5.9)$$

where \hat{V} is the model visibility at u_i, v_i and θ are the model parameters. While closure products are immune to gains their statistical properties are more complicated. Namely, the likelihoods for closure phases and amplitudes are markedly non-Gaussian at low signal-to-noise-ratio (SNR) (Thompson et al., 2017; Broderick et al., 2020). As a result low SNR data are usually flagged, and could potentially bias results.

The downside of fitting complex visibilities is that the complex gains have to be forward modeled. This could potentially introduce a large number of additional parameters in the model (~ 250 for the M 87 observations). To combat this, we will use THEMIS' complex gain marginalization scheme. This scheme uses a Laplace approximation to approximate the marginalization using a saddle-point approximation at each MCMC step (see Appendix A for a review). Intuitively, this effectively acts as a self-calibration for each proposed image during sampling. In practice, we find that this adds a small additional cost for each likelihood evaluation, while greatly lowering sampling time compared to directly fit the gains.

Finally, we follow the data processing choices used in EHTC VI. That is, we will be using EHT data that has been coherently averaged over scans. By using scan-averaged data we are effectively assuming that the gains are constant over each scan, which is approximately true for M 87 (EHTC III). Additionally, we add an 1% fractional visibility error to the reported thermal noise:

$$\sigma_{ij}^2 \rightarrow \sigma_{ij}^2 + (0.01V_{ij})^2. \quad (5.10)$$

This models the polarization leakage observed for M 87.

To create simulated data for THEMIS we used the `eht-imaging` and `THEMISPy` packages following the same procedures as [EHTC VI](#) and [Chapter 3](#). Namely, we added a zero baseline unresolved flux to bring the total flux up M 87’s 1.2Jy, station-based complex gains, and the effects of a 1% polarization leakage non-closing error. Finally, we note that the EHT measures the complex visibilities in two frequencies, commonly called HI and LO band. In [Section 5.4](#) we only consider the LO band datasets. For [Section 5.5](#), we use both the LO band and HI+LO band datasets.

5.4 Geometric Tests

In principle, Bayesian imaging should be immune to the ellipticity bias in the `eht-imaging` top set. To test this, we will consider the same ellipticity validation from [Section 3.3](#). Namely, we consider an elliptical Gaussian ring with ellipticity $\tau = 0.187$ at various orientations. Given the timescale of Bayesian imaging we only consider the orientations $\xi_\tau = -45^\circ, 0^\circ, 45^\circ, 90^\circ$, since the remaining cases just flip the location of the brightness maximum.

To validate Bayesian imaging, we broke the analysis into two steps. First, we attempted to find the optimal number of pixels given the observed data. Given the computational timescale of Bayesian Imaging, we only attempted this on the elliptical ring with orientation 90° . We chose this orientation since this was the case where the `eht-imaging` top set had the largest biases (see [Figure 3.3](#)). Second, once the optimal number of pixels was identified, we fixed the number of pixels and imaged the other orientations i.e., $\xi_\tau = \pm 45^\circ$, and 0° .

5.4.1 Pixel Optimization

To find the optimal raster grid we used the pixel configurations: 6x4, 5x5, 6x6, and 8x8. This choice was inspired by [Broderick et al. \(2020\)](#), which found that the optimal number of pixels was a 5×5 raster for 5 GRMHD simulations using the uv coverage on April 11. Using this as our starting point, we also considered two higher resolutions to test whether increased resolution drastically changes the evidence and image structure. We also considered the 6×4 resolution since this roughly matches the true ellipticity of the ring.

To sample from the posteriors we used THEMIS’ new sampling interface described in the previous chapter. Furthermore, the sampler settings were inspired by the validation exercise in [Section 4.6](#). For the local posterior exploration, we used the Stan NUTS sampler with

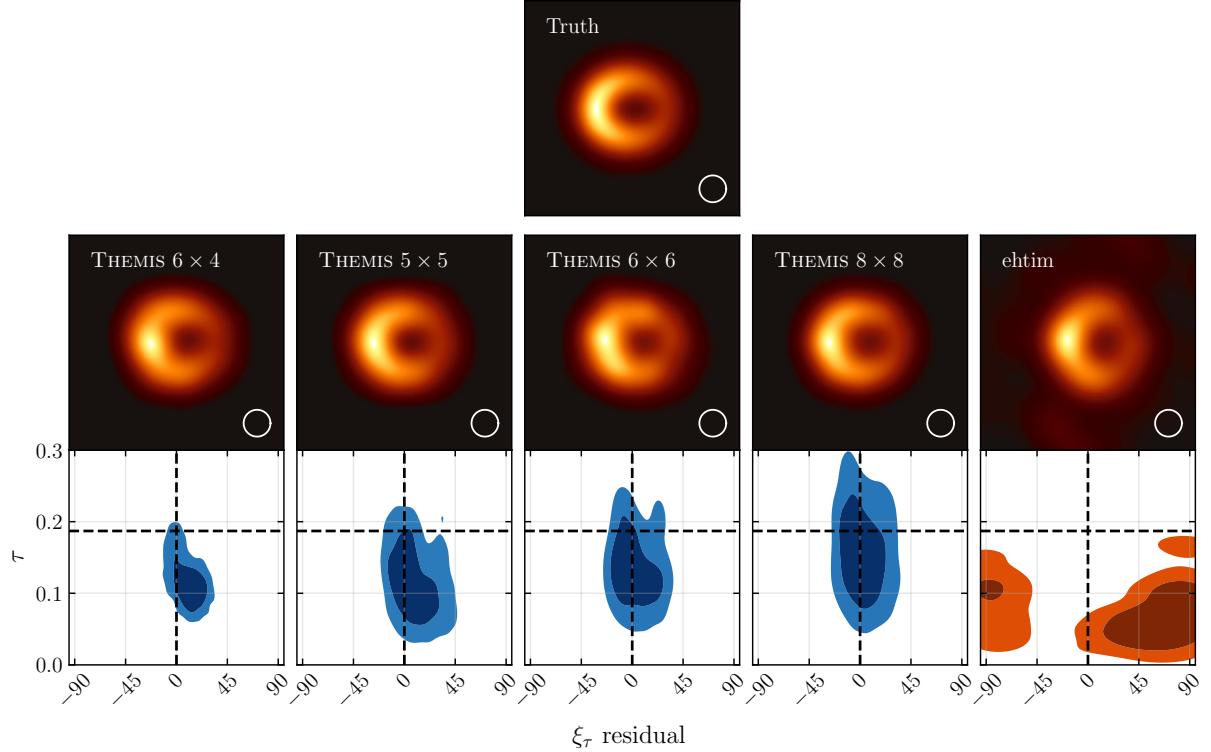


Figure 5.1: Comparison of the mean image reconstructions (middle row) and ellipticity posterior (bottom row) compared to the sky-truth image (top row). The probability contours show the 68% and 95% probability regions. All images have been blurred with a circular Gaussian with a $15\ \mu\text{as}$ FWHM. The first row columns are the results using the different number of pixels described in the text. The last columns show the results from the `eht-imaging` top set. Overall we find the ellipticity is within the 95% probability regions for all Bayesian imaging techniques. This is the opposite of the `eht-imaging` top set, which failed to produce a single image representing the true ellipticity and orientation.

a diagonal mass matrix and 2000 adaptation steps. To enable global posterior exploration, we used the DEO parallel tempering sampler with 60 tempering levels. The PT sampler was run for 9–10 adaptation rounds, where we used the Stan adaptation scheme at the beginning of each round. We found an optimal global communication barrier $\Lambda \approx 15 – 16$, implying an optimal round-trip rate of 0.032. After 9 rounds the ladder had a communication barrier of $E \approx 20 – 21$, which gives a round-trip rate of 0.023. All chains were run until the split- $\hat{R} < 1.1$ for all image raster c_{ij} parameters.

Table 5.1: Elliptical Ring Raster Survey Results

	8×8	6×6	5×5	6×4
ΔBIC	240.6	68.9	5.2	0.0
χ^2_{red}	1.49	1.28	1.23	1.22

The mean images from each run are shown in the second row of [Figure 5.1](#). For the 6×4 , 5×5 , and 6×6 rasters, the posterior was very multi-modal, usually displaying 3-4 distinct image clusters.

To move from the posterior of c_{ij} to image features, we randomly selected 2000 images from the last tempering round. These images were then run through VIDA using the templates described in [Section 5.2](#). The resulting ellipticity and orientation for each raster resolution are shown in [Figure 5.1](#). We also include the `eht-imaging` results from [Chapter 3](#) in the right-most column. We find that the true ellipticity and orientation are within the 95% confidence regions in all cases. This contrasts the `eht-imaging` topset, which, as detailed in [Chapter 3](#) does not contain a single reconstruction consistent with the on-sky image.

The impact of adding more pixels/model freedom can be seen moving from left to right. As the number of image pixels increases, the posterior tends to become broader. This is expected since there the larger number of parameters enables the model more degrees of freedom. This pattern could be continued further. Namely, we could keep adding more and more pixels. As we do, the posteriors would become broader until we are effectively sampling from the prior. However, in this case the marginal likelihood would be very low.

To find the preferred raster grid we turn to the Bayesian-information-criteria (BIC). The BIC is defined as

$$\text{BIC} = \chi^2 + k \ln(N), \quad (5.11)$$

where k is the total number of non-gain parameters fit, N is the number of data points, and χ^2 is the minimum chi-square for the non-gain marginalized likelihood, i.e., [Equation 5.9](#)². The BIC is a simple approximation Bayesian evidence (see [Appendix B](#) for a review) and compares the optimal performance of the model (first term in [Equation 5.11](#)) to its complexity (second term). The best model is the one that minimizes the BIC or conversely approximately maximizes the Bayesian evidence.

The results of the parameter survey are shown in [Table 5.1](#). Overall we find that the BIC and reduced chi-square are the smallest for the 5×5 and 6×4 models. These results

²Note we don't include gains, since they cancel when finding the ΔBIC .

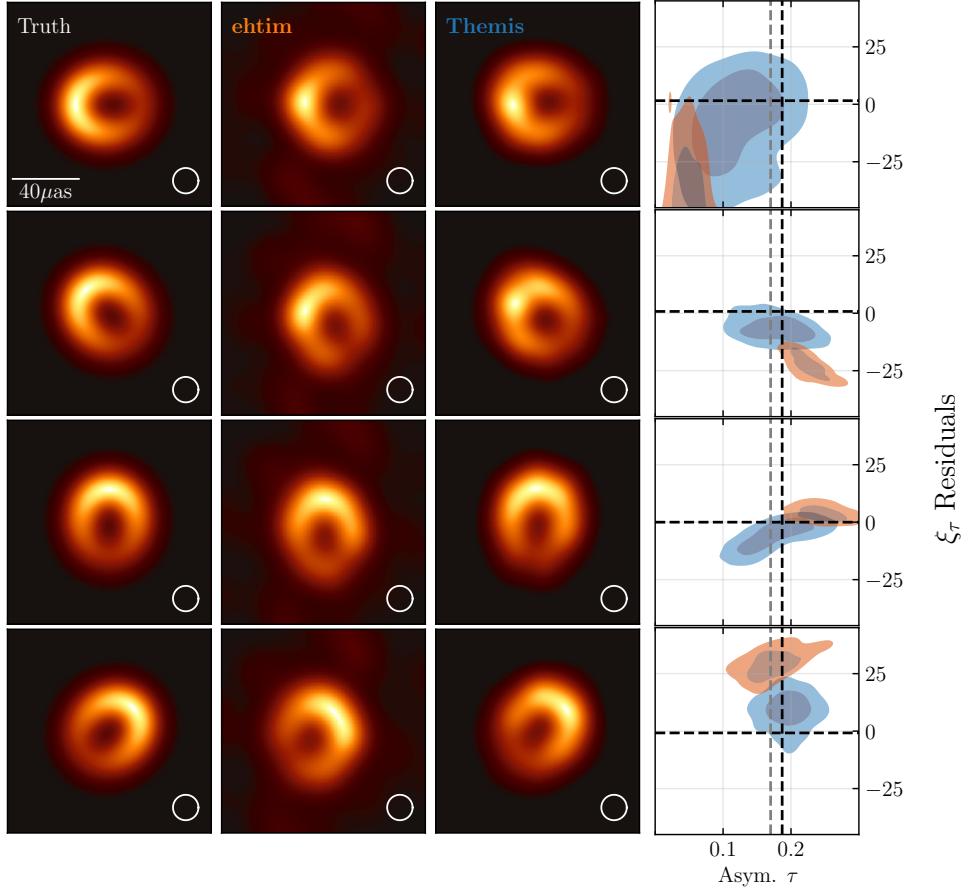


Figure 5.2: Comparison between `eht-imaging` and `THEMIS`' Bayesian image reconstructions of the elliptical ring test from Chapter 3. The left column shows the on-sky truth image. The second columns shows the mean image from the `eht-imaging` topset. The third column shows the mean image from the `THEMIS` posterior using a 5×5 raster. All images have been blurred by a Gaussian with a $15 \mu\text{as}$ FWHM. The rightmost column shows the joint distribution for the recovered ellipticity, τ , and its position angle, ξ_τ for `eht-imaging` (orange) and `THEMIS` (blue). The contours display the 68% and 95% probability regions. The black dashed line shows the on-sky truth values, and the grey dashed line the truth values after being blurred by a Gaussian with $15 \mu\text{as}$ FWHM.

are consistent with the results from Broderick et al. (2020). Namely, Broderick et al. (2020) found that the BIC favored the 5×5 grid for simulated data from GRMHD models given in Event Horizon Telescope Collaboration et al. (2019e). Additionally, unlike the 6×4

resolution, the 5×5 raster does not presume an ellipticity source structure. Motivated by this reasoning, we use the 5×5 raster in all future experiments.

5.4.2 Orientation Dependence of Bayesian Imaging

Given that 5×5 optimizes the BIC, we now repeat our analysis on the simulated elliptical ring data with $\xi_\tau = 45^\circ, 0^\circ, -45^\circ$. We used the same sampler setting as above, and ran for a similar number of steps. To construct the feature posteriors, we again randomly selected 2000 images from the last tempering rounds chain and extracted the image features using VIDA. The mean images for THEMIS and `eht-imaging` are shown in [Figure 5.2](#).

Visually from [Figure 5.2](#) we see that THEMIS' mean image appears more similar to the ground truth image. Namely, it does not appear to have the same N-S bias as the `eht-imaging` top set. Quantitatively we find that THEMIS's ellipticity posterior contains the truth in its 95% contours for all the orientations considered. Interestingly, we found that for the $\xi_\tau = 90$ orientation THEMIS reported substantially larger uncertainty, and has regions with similar values as the `eht-imaging` top set. This implies that the M 87 uv coverage is causing increased uncertainty for on-sky images with this ellipticity orientation. However, it also demonstrates that Bayesian imaging is immune to the failures of the M 87 top set. Namely, in all cases the truth is contained in the posteriors.

5.5 M 87 Imaging

The previous section demonstrated that Bayesian imaging can recover the true ellipticity. This result demonstrates that we can directly interpret the measured ellipticity as representing the ellipticity of the on-sky image. Therefore, in this section we will apply the Bayesian imaging and VIDA pipeline to the 2017 EHT M 87 data.

To analyze the 2017 M 87 EHT data, we will consider two separate data sets. First, to compare the results with those in [Chapter 3](#), we will consider the LO band only data. Second, we will combine HI and LO bands and fit the total HI+LO band data. We expect that the combined data will increase the constraining power since we are effectively doubling the amount of data. Additionally, by splitting the data into two sets, we are effectively performing a jackknife test. Namely, if our model is an adequate representation of the on-sky image, we expect consistent results between the two datasets.

For both datasets, we will use the same image model. Namely, following the parameter exploration above, we will consider a 5×5 raster with adaptive field of view and orientation.

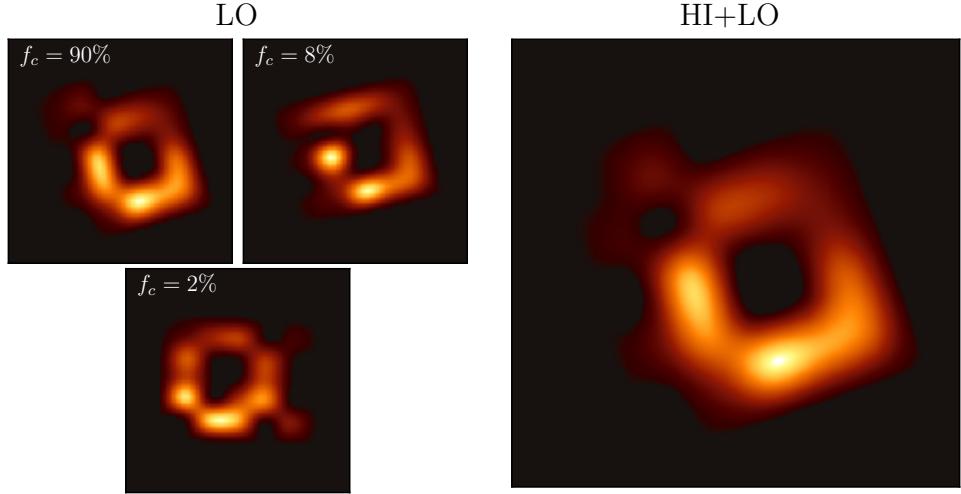


Figure 5.3: Example image reconstructions of M 87 LO band on April 11 (left 3 plots) and HI+LO band (right). For LO band we show three images corresponding to the three image clusters seen in the MCMC chains. In the top left corner of the LO band plots, we show the percentage of images in each cluster taken from the last 50% of the MCMC chain. For HI+LO band we find that the extra data removes the two lower percentage modes.

Note that for the combined HI+LO analysis, we do not assume that the gains are the same across bands. Additionally we apply THEMIS' gain marginalization scheme separately to each band.

For the LO band data, we used similar sampler settings for the elliptical ring simulated data. Namely, 60 tempering levels, and 2000 adaptation steps at the beginning of each tempering round. All samplers were run until the split- \hat{R} was less than 1.03. On April 5, 10, and 11 we found that the posterior was multi-modal, while on April 6 we only found a single mode.

Each image mode on April 5, 10, and 11 had a slightly different image structure. To identify the image clusters we used a k-means clustering algorithm on the last 50% of the chain from the last tempering round. The mean images of each cluster and their relative fractions on April 11 are shown in [Figure 5.3](#). This demonstrates the power of Bayesian imaging and our sampling techniques. We were able to identify three dominant image modes, and assign a probability to each mode.

For the combined HI and LO data, we used 320 tempering levels. Each chain was run for > 50000 MCMC steps and achieved a split- $\hat{R} \leq 1.03$. Interestingly, we found that the

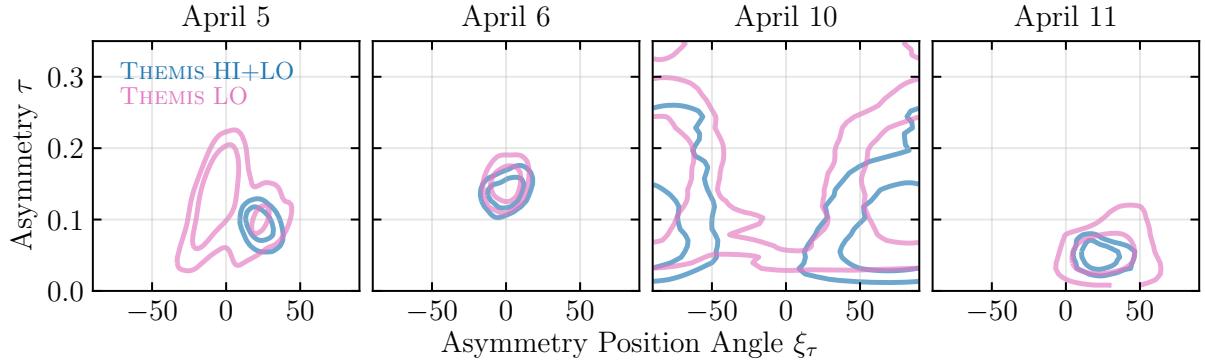


Figure 5.4: Joint ellipticity value and orientation from M 87 on April 5, 6, 10, and 11. The contours show the 68% and 95% probability regions from THEMIS using the combined HI+LO band data (blue), LO band only (pink). The LO band and HI+LO ellipticity results are consistent between bands.

additional data from fitting the combined bands eliminated the multi-modality on April 5 and 11. This is shown in [Figure 5.3](#) for April 11. We see that while the LO band only fits have three distinct modes, once we include the HI band data, the other two lower probability modes disappear. On April 10, we found the same three modes for the HI+LO band as the LO band only fits.

To create the feature posteriors, we created 2000 images from the last tempering round’s chain and ran VIDA on it. The resulting ellipticity and orientation joint posteriors are shown in [Figure 5.4](#). Overall we find that combined HI+LO band ellipticity posterior is contained in the LO band posterior. This is the expected result for a properly specified model, where the increase in data should cause the posterior to consistently shrink. Given this result, we will only consider the combined HI+LO band results in the following discussion since they provide the strongest constraint.

In [Figure 5.5](#) we compare the `eht-imaging` and THEMIS results across each observation day. The top two rows show the mean THEMIS and `eht-imaging` images blurred by a Gaussian with $15\mu\text{as}$ FWHM. The THEMIS and `eht-imaging` results are discrepant at over 2σ on April 5, 6, and 11. On April 10 the ellipticity and position angles do overlap. However April 10 has substantially less data than the other days. As a result the relative error bars are much larger, likely aiding in the consistency.

Focusing on the THEMIS results, we see that on April 5 and 6 the resulting ellipticity distributions are consistent at a 2σ level, while April 10 and 11 are consistent with each other. Comparing April 5/6 with April 10/11, we see that there is no significant evidence

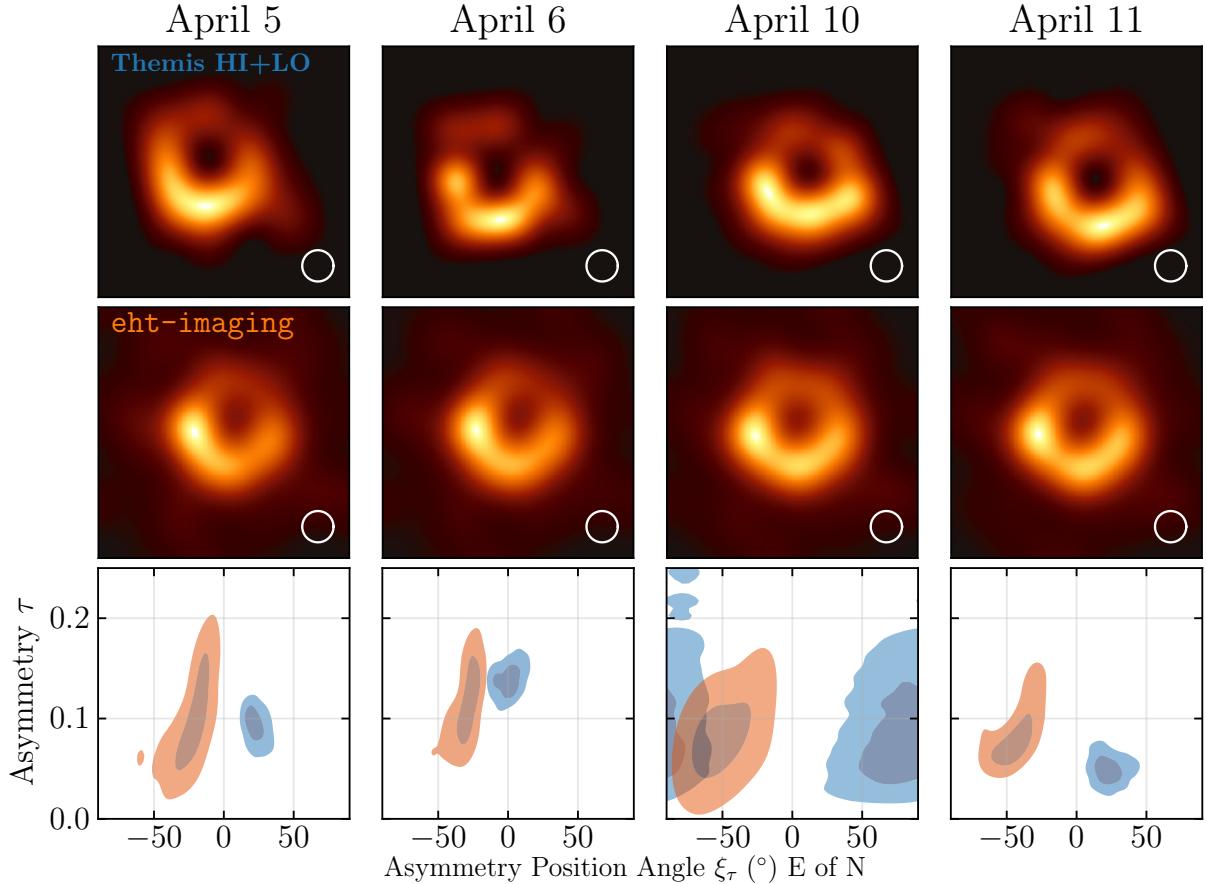


Figure 5.5: Mean images from the THEMIS Bayesian imaging posterior (top) and ehtim topset (middle) both blurred by a $15 \mu\text{as}$ FWHM Gaussian. The bottom row shows the joint distribution of the image ellipticity vs position angle for the THEMIS results(blue) and **eht-imaging** results (orange)

($> 2\sigma$) for ellipticity evolution M 87 across the 2017 observations.

5.5.1 Interpreting the M 87 ellipticity

In [Section 5.4](#), we demonstrated that Bayesian imaging gives a faithful reconstruction of ellipticity. This means that the calibration procedure applied in [Section 3.4](#) is not needed. Additionally, since we do not need to use GRMHD simulations to calibrate for imaging

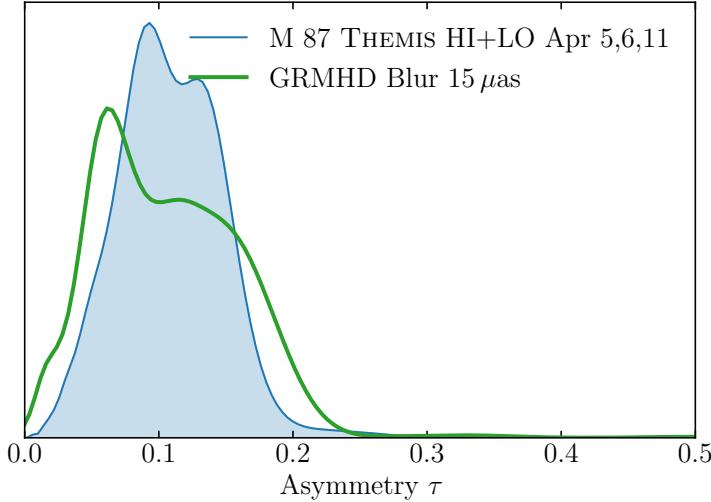


Figure 5.6: Comparison of recovered ellipticity of M 87 using Bayesian imaging (blue), and the expected ellipticity from GRMHD simulations blurred with a $15 \mu\text{as}$ Gaussian kernel (green line). Both the GRMHD and Bayesian imaging asymmetries were found using VIDA. We find that the recovered M 87 ellipticity is within the expected GRMHD ellipticity distribution and thus is consistent with theoretical expectations.

uncertainty, we can now ask whether the Bayesian imaging results are consistent with the expectation from GRMHD simulations.

A thorough analysis of this consistency question would require imaging a large number of GRMHD simulations using the Bayesian imaging pipeline from [Section 5.5](#). However, given the computational timescale ($\sim 5\text{-}6$ days) of Bayesian imaging, this is not feasible. However, from ([EHTC VI](#)), the image structure uncertainty will be dominated by accretion uncertainty. Therefore, to check whether the measured ellipticity is consistent with GRMHD simulations, we used the marginal ellipticity posterior from [Section 3.4.1](#). Recall, that this was measured by applying VIDA to the raw GRMHD images, blurred with a $15 \mu\text{as}$ FWHM Gaussian. This accounts for the expected ellipticity from accretion turbulence around M 87, and the blurring models the finite resolution of the current EHT array. [Figure 5.6](#) compares the THEMIS ellipticity posteriors to the distribution of expected GRMHD ellipticity. We see that the Bayesian imaging posterior over the combined days (sans April 10) are consistent with the expected ellipticity from simulations. This result implies that GRMHD simulations are consistent with the M 87 ellipticity reconstructions.

5.6 Conclusion

We have demonstrated that Bayesian imaging can reliably recover the on-sky ellipticity. The ellipticity posterior is robust to the number of pixels chosen for the raster and ellipticity orientation. This result is a demonstration of the power of Bayesian imaging. While it is computationally more expensive than traditional RML methods, it can reliably quantify image features. Furthermore, we did not need to use any calibration procedures to correct for parameter bias.

The Bayesian ellipticity posteriors from M 87 are robust to coverage, and can measure the actual ellipticity of the on-sky source. Analyzing the results from [Figure 5.5](#), we see that the marginal ellipticity posterior satisfies $\tau \lesssim 0.15$ on all days except April 10. However, April 10 has substantially worse coverage and even then only extends up to $\tau = 0.25$. Therefore, using the April 5,6, and 11 results, we see that Bayesian imaging improves the ellipticity constraint by a factor of 3.3 compared the the calibrated M 87 top set results. Furthermore, we demonstrated that M 87's ellipticity posteriors are consistent with the expectations from GRMHD simulations and thus are consistent with general relativity. In the future, this ellipticity constraint can be further improved by fitting the entire 2017 M 87 dataset.

Unfortunately, this ellipticity constraint does not tell us much about the nature of spacetime near the horizon. It is possible for non-GR metrics to have similar amounts of ellipticity ([Johannsen et al., 2016](#)). Part of the problem is that we are only looking for non-GR signatures from static signatures that require high resolution. One way to improve this is through dynamics ([Broderick et al., 2021](#); [Moriyama et al., 2019](#)). Unfortunately M 87, given its mass, is approximately static over the 2017 observations. However, the other primary EHT source Sgr A*, is a ~ 1000 times lighter and thus changes on the timescales of seconds. In the next chapter, we will propose how to use Sgr A*'s strong-variability to test GR by tomographically mapping spacetime using orbiting hot spots.

Chapter 6

Space Time Tomography with the Event Horizon Telescope

6.1 Introduction

One of the primary science goals of the EHT is to probe accretion and spacetime in the strong-gravity regime. In the previous chapters we explored the static image properties of M 87. Static imaging, however, creates a degeneracy between when and where emission arises near the black hole. By exploring time variability, we can break this degeneracy and explore the dynamics of the emission region and structure of spacetime.

While M87 is static over a day, Sgr A* is constantly variable. Furthermore, Sgr A* displays strong variability through flaring events (Genzel et al., 2003) that can often (especially for bright X-ray flares) be seen (almost) simultaneously across multiple bands (Fazio et al., 2018), from sub-millimeter (Fish et al., 2011), to infrared (Gillessen et al., 2006; Witzel et al., 2012; Witzel et al., 2018) and X-ray (Neilsen et al., 2013; Ponti et al., 2017). This emission appears to come from a compact region near the innermost stable circular orbit (ISCO) of the black hole and is presumed to be from dynamical structures within the accretion flow (Marrone, 2006; Gillessen et al., 2006; Gravity Collaboration et al., 2018). An explanation for these flares comes from the creation of localized “hot spots” of non-thermal electrons in the accretion disk surrounding the black hole and has been proposed by several authors: Broderick & Loeb (2005, 2006); Eckart et al. (2006) and previously Dovčiak et al. (2004). A natural origin is magnetic-reconnection events within the accretion disk analogous to solar flares, an unavoidable consequence of radiatively inefficient accretion models.

Previous work on modeling orbiting hot spots assumed compact spherical Gaussian structures that remain coherent during its orbit(Broderick & Loeb, 2005, 2006). In general, however, hot spots are expected to be embedded within a differentially rotating accretion disk, and therefore will shear and expand. Furthermore, shearing due the differential flow of an accretion disk can lead to large observational differences for NIR flares (Eckart et al., 2008a, 2009; Zamaninasab et al., 2010). For the EHT, including shearing may be imperative since it is sensitive to horizon scale physics. To address these concerns, we developed a computationally efficient model including generic shear and expansion, while ensuring that spot number density is locally conserved.

As hot spots orbit, they probe different parts of spacetime. By observing a flare, we not only probe the null structure of spacetime from the emission but also how massive matter evolves in the vicinity of the event horizon. As we will show below, observing a single hot spot with the EHT may lead to high precision spin measurements. Furthermore, since each hot spot will form at a different radius, every flare will probe different regions of spacetime. Combining multiple flares would then amount to constructing a tomographical map of spacetime, leading to a new test of GR in the vicinity of black holes.

In practice, recovering the hot spots from EHT observations could be difficult. The effective beam size of the EHT is $13\mu\text{as}$, meaning that the hot spot may not be sufficiently resolved by the EHT to precisely probe spacetime. Furthermore, there are several important systematics present for Sgr A*, such as scattering and the background accretion flow. To address these questions we used **THEMIS** (Broderick et al., 2020), a Bayesian parameter estimation framework designed for use with the EHT. Therefore, we will numerically explore the ability of the EHT to perform inference on hot spots using synthetic data that matches the configuration of the EHT 2017 array and address the impact of some potential systematics.

This chapter is organized as follows. In [Section 6.2](#) we present an original hot spot model that incorporates shearing and expansion while conserving particle number. [Section 6.3](#), explores the ability of the EHT in the 2017 configuration, to extract a hot spot from potential observations of Sgr A*. As a result, we analyze whether the differential flow parameters such as angular velocity and black hole spin, are degenerate. Additionally, we study how the background flow, scattering, and the accretion disk inclination relative to our line of sight impact our results. [Section 6.4](#) details how multiple hot spots can be used to tomographically map spacetime using the EHT and constructs hypothetical maps using the EHT.

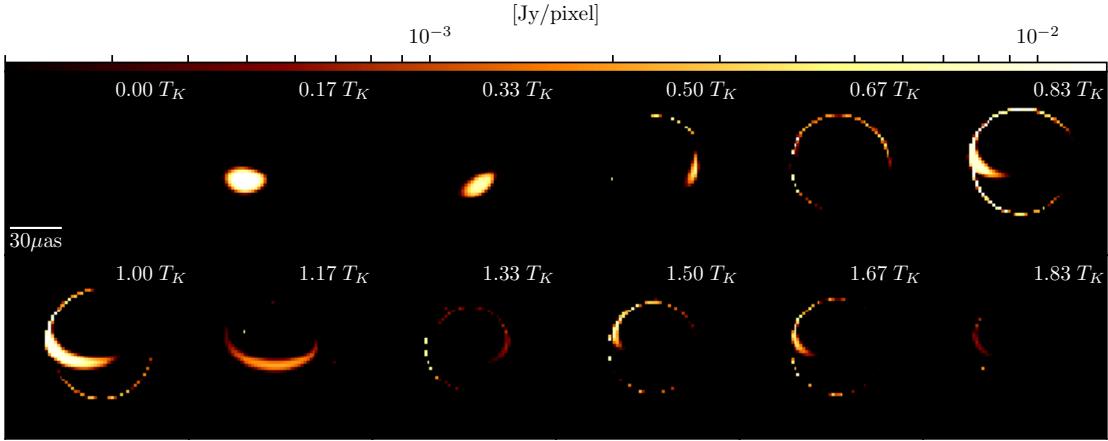


Figure 6.1: Frames of a shearing spot movie intensity map in log-scale, with 64×64 image resolution set around Sgr A*. The movie lasts for two Keplerian orbits at a radius of $5.23M$, and contains 12 frames in total. The parameters are $a_* = 0.5$, $\cos \Theta = 0.5$, $n_0 = 5.5 \times 10^7$, $R_s = 0.5M$, $r_0 = 5.23M \simeq 1.25 r_{\text{ISCO}}$, $\phi_0 = -90^\circ$, $\alpha = 0.05$, $\kappa = 0.95$.

6.2 Hot spot model

Hot spot models have been used to explain flares and variability in Sgr A* observations using coherent Gaussian hot spots (Broderick & Loeb, 2005, 2006). Later models allowed for adiabatic expansion (Eckart et al., 2009) and shearing (Zamaninasab et al., 2010). However, in the latter work, the emission was restricted to a 2D disk and radiative transfer effects were ignored. Here, we describe a model that includes both shearing and expansion of a hot spot along a stationary and axisymmetric velocity field. Furthermore, this model includes the 3-dimensional structure of the hot spot and includes the effects of radiative transfer as will be described below. Additionally, we describe a semi-analytical procedure for evolving the hot spot density that is marginally more computationally expensive than the coherent hot spot model.

6.2.1 Density profile evolution

In Broderick & Loeb (2005, 2006), hot spots are modeled by orbiting, symmetric Gaussian electron overdensities. The orbital position of the center of the spot, y_0^μ is determined by integrating the accretion flow four-vector field u^μ around which the hot spot number

density is given by,

$$n_e(y^\mu) = n_0 e^{-(\Delta r^\mu \Delta r_\mu + (\Delta r^\mu u_\mu)^2)/(2R_s^2)}, \quad (6.1)$$

where $\Delta r^\mu = y^\mu - y_0^\mu$, is the displacement vector from the center of the Gaussian spot, R_s the spot size, and u_μ is evaluated at the spot center. While this model is computationally efficient, it ignores the potential for differential motion within the spot. For a Keplerian velocity field, this approximation is rapidly violated: the inner edge of the spot has advanced relative to the outer-edge by one radian in $r/3\pi R_s$ orbits; within $1/3\pi$ orbits it has advanced by R_s .

Therefore, we will assume that a hot spot will travel passively on some specified (e.g. background accretion flow) velocity field u^μ . That is, the density is described by the continuity equation coupled with the condition that the particles move with the background flow:

$$\nabla_\mu(n_e u^\mu) = 0, \quad (6.2)$$

$$\frac{dx^\mu}{d\tau} = u^\mu. \quad (6.3)$$

We will denote the solution family of [Equation 6.3](#) by the map $\varphi_\tau(y^\mu)$. By fixing y^μ we can consider $x^\mu(\tau) = \varphi_\tau(y^\mu)$ as describing the motion of a particle at some time τ . Namely, $x^\mu(\tau)$ satisfies [Equation 6.3](#), with initial condition $x^\mu(\tau = 0) = y^\mu$. On the other hand, if we consider τ as fixed then $\phi(y^\mu) = \varphi_\tau(y^\mu)$ describes the coordinates of a family of observers at rest with the background flow. Therefore, we can say for small τ that $\varphi_\tau(y^\mu)$ forms a 1-parameter family of diffeomorphisms. Therefore, we can solve [Equation 6.2](#) using the method of characteristics, giving

$$u^\mu \partial_\mu n_e + n_e \nabla_\mu u^\mu = \frac{d}{d\tau} n_e + n_e \nabla_\mu u^\mu = 0 \quad (6.4)$$

and

$$n_e(\tau, x^\mu) = n_{e,0}(y^\mu) \exp\left(-\int_{\tau_0}^{\tau} \nabla_\mu u^\mu d\tau\right), \quad (6.5)$$

where $y^\mu = \varphi_\tau^{-1}(x^\mu)$ is the initial position of the spot and $n_{e,0}$ the initial proper density profile. We can simplify this further by noting

$$\nabla_\mu u^\mu = \frac{1}{\sqrt{-g}} \partial_\mu(\sqrt{-g} u^\mu), \quad (6.6)$$

and thus,

$$n_e(\tau, x^\mu) = n_{e,0}(y^\mu) \frac{\sqrt{-g(y^\mu)}}{\sqrt{-g(x^\mu)}} \exp\left(-\int_{\tau_0}^{\tau} \partial_\mu u^\mu d\tau\right). \quad (6.7)$$

Interpreting this result physically, we see that there are two forms of expansion included in the model. The first is spacetime expansion and is encoded by the ratio of the metrics. The second is the expansion of the velocity field itself irrespective of the background spacetime. Note that for a hot spot outside the innermost stable circular orbit (ISCO) traveling on a Keplerian orbit, [Equation 6.12](#) simplifies to $n_e(\tau, x^\mu) = n_{e,0}(y^\mu)$, since $g(x^\mu)$ is constant and $\partial_\mu u^\mu$ vanishes. This does not mean that there is no deformation, as $y^\mu = \varphi_\tau^{-1}(x^\mu)$. Instead, the only deformation is due to shearing which does not affect the proper density¹. Note however, that this is an idealization and in reality the hot spot will adiabatically expand during its evolution, meaning that its proper density will decrease over time.

In this chapter, we will focus on hot spots orbiting in the equatorial plane, although the density extends outside. Furthermore, we will assume that the spot has negligible vertical motion compared to radial and azimuthal motion. To describe our vector field we follow [Pu et al. \(2016\)](#). A form of the accretion flow that obeys the restrictions mentioned above is

$$u^\mu = (u^t, u^r, 0, u^t\Omega), \quad \Omega = u^\phi/u^t. \quad (6.8)$$

The normalization condition $u^\mu u_\mu = -1$ for a black hole metric in Boyer-Lindquist like coordinates gives

$$u^t = \sqrt{\frac{1 + g_{rr}(u^r)^2}{-g_{tt} - 2\Omega g_{t\phi} - \Omega^2 g_{\phi\phi}}}. \quad (6.9)$$

From this we can see that we require that $g_{tt} + 2\Omega g_{t\phi} + \Omega^2 g_{\phi\phi} < 0$. To specify u^r and Ω we will use a combination of Keplerian and free-fall motion:

$$u^r = u_K^r + \alpha(u_{\text{ff}}^r - u_K^r), \quad (6.10)$$

$$\Omega = \Omega_K + (1 - \kappa)(\Omega_{\text{ff}} - \Omega_K), \quad (6.11)$$

where $\alpha, \kappa \in [0, 1]$, are two free-parameters that control the rate of free-fall and the sub-Keplerian motion respectively. Note that our definition of α differs from that of [Pu et al. \(2016\)](#). For the Keplerian component, outside the ISCO, $u_K^r = 0$. Inside the ISCO $u_K^r \neq 0$ and is specified by matching the energy and angular momentum at the ISCO. This choice of velocity field brackets a useful collection of accretion flows. For example, taking $\alpha = 0, \kappa = 1$ gives a Keplerian orbit and $\alpha = 0, \kappa = 0$ free-fall motion.

Due to the fact that this vector field is independent of the coordinates t and ϕ , we have

¹We define shearing as the symmetric traceless part of the tensor $\nabla_\mu u_\nu$. Since it is trace-free does not directly impact the proper density of the hot spot (see [Equation 6.7](#)).

that $\partial_\mu u^\mu = \partial_r u^r$. Therefore, $\partial_\mu u^\mu = (u^r)^{-1} \dot{u}^r$, and [Equation 6.5](#) simplifies to

$$n_e(\tau, x^\mu) = n_{e,0}(y^\mu) \frac{\sqrt{-g(y^\mu)}}{\sqrt{-g(x^\mu)}} \frac{u^r(y^\mu)}{u^r(x^\mu)}. \quad (6.12)$$

This semi-analytic formula greatly increases the computational speed of the hot spot, and is the same order of computational complexity as the coherent spot used in [Broderick & Loeb \(2005\)](#).

As a final note, in principle, any smooth function could be used for the initial density profile. However, in this chapter, we will assume that the spot is initially given by [Equation 6.1](#). There are two reasons for this. First, this profile allows us to compare the hot spot evolution results to [Broderick & Loeb \(2005, 2006\)](#). Second, since any image will be distorted by the interstellar scattering screen ([Bower et al., 2006; Johnson et al., 2018](#)), whose diffractive or blurring component is effectively a Gaussian with a semi-major axis of $22\mu\text{as}$ most small-scale structure of the hot spot will be unresolved. Therefore, since we are assuming that a hot spot forms from local microphysics, i.e. fast magnetic reconnection, we expect it to be contiguous, and the initial profile can be approximated as a Gaussian.

6.2.2 Radiative transfer and Ray-tracing

To create hot spot models that can be compared to EHT data, we need to create realistic images. This means that near the black hole, general relativistic (GR) and radiative transfer effects need to be included. These effects include the geometric and gravitational time delays across the source (often called “slow light”) and the strong gravitational lensing that magnifies the emission region and produces secondary images associated with photons that complete half orbits around the black hole. Furthermore, optical depth becomes important near the black hole since material moving towards the detector will have an increased apparent density due to Doppler effects, making it optically thick.

To incorporate these effects we use the covariant ray-tracing and radiative transfer code VRT² (vacuum ray-tracing radiative transfer). For the EHT observation band (230GHz) we assume the hot spot spectra in the plasma rest frame is given by the synchrotron self-absorption model from [Broderick & Blandford \(2004\)](#), with a local plasma energy spectral index of $s = 2.25$. At the observing frequency of the EHT ($\sim 230\text{GHz}$), synchrotron cooling processes will be sub-dominant to the shearing timescale. The shearing timescale is roughly the orbital period of the hot spot around Sgr A*. Taking the hot spot to be around the ISCO, we get $t_{\text{shear}} \sim 10 - 30$ min for a $4 \times 10^6 M_\odot$ black hole. The timescale for synchrotron cooling can be estimated from $t_{\text{synch}} \sim 3 \times 10^7 \nu_9^{-0.5} B^{-3/2}$ s, where ν_9 is the frequency in

GHz and B is in Gauss. Taking $B \sim 10 - 50$ G we find that $t_{\text{synch}} \sim 1 - 20$ hours at 230 GHz. Therefore, in this chapter we ignore the cooling break and evolution during the hot spots orbit. Note that in the other bands, e.g., the near-infrared and X-ray, cooling and inverse Compton effects likely become important (Fazio et al., 2018) and will need to be included.

To model the magnetic field assumed to arise from an accretion disk, we followed Broderick et al. (2016) and used a toroidal magnetic field with a fixed plasma beta set to 10. Below we will also consider what happens when the hot spot is embedded in an accretion flow. In this case, we use the accretion flow model from Broderick et al. (2016). This model includes thermal electrons set in a radial power law in density and temperature, where the power-law indices for the thermal electron and temperature distribution are $-1.1, -0.84$ respectively. Furthermore, we also include non-thermal electrons with a radial power-law index -2.02 for the number density, and 1.24 for the photon spectral index. Note that these parameters were chosen to match the values in Broderick et al. (2016), which are the best fit values to the spectrum of Sgr A*.

The optical depth of the hot spot depends on a number of quantities, such as the proper density of the hot spot, its orbital parameters, and the orientation of the orbital plane relative to our line of sight. For instance, a hot spot with a 1 Jy flux near the ISCO at 230 GHz with an inclination angle of 60° , will tend to appear to be optically thick when it is moving towards the observer due to Doppler beaming. Furthermore, as the hot spot shears its effective optical depth will change. All of these effects are automatically included into the relativistic radiative transfer that occurs in the construction of each frame of the movie.

This completely describes the shearing hot spot model used in this chapter. In summary, this model, ignoring any spectral information, requires 10 parameters to describe the evolution which are:

1. spin parameter a_* .
2. cosine of the inclination, $\cos \Theta$ of the black hole relative to the image screen.
3. Spot electron density n_0 in Equation 6.1.
4. Spot characteristic size, R_s , in units of M^2 from Equation 6.1.
5. Spot injection time, t_0 , in units of M for an observer: the time which the spot is instantaneously injected into the accretion flow. The actual spot appears at a fixed

²We are using geometrical units, where $G = c = 1$ in this chapter.

proper time for an observer in a locally flat co-moving frame with u^μ . This means for the observers time, the hot spot gradually starts to appear.³

6. Initial hot spot radius, r_0 , i.e. the position of the spot center in Boyer–Lindquist coordinates when it is initially injected.
7. Initial hot spot azimuthal angle, ϕ_0 , i.e. the angle in Boyer-Lindquist coordinates of the hot spot center when it is injected into the accretion flow.
8. Radial accretion flow parameter, α , in [Equation 6.10](#).
9. Angular accretion flow parameter, κ , in [Equation 6.11](#).
10. Position angle of black hole spin and orbital axis, ξ .

[Figure 6.1](#) presents a 12 frame movie with the parameters $a_* = 0.5$, $\cos \Theta = 0.5$, $n_0 = 5.5 \times 10^7$, $R_s = 0.5$, $t_0 = -6M$, $r_0 = 5.23M \simeq 1.25r_{\text{ISCO}}$, $\phi_0 = -90^\circ$, $\alpha = 0.05$, $\kappa = 0.99$, $\xi = 0^\circ$. The initial time of the spot was chosen to be $-6M$ due to time delay effects from ray-tracing. The inclination was chosen to be equal to the expected inclination for a uniform distribution on a sphere, which is close to the observed inclination found in [Broderick et al. \(2016\)](#). Different inclinations will be explored in [Section 6.3.4](#). Additionally, we chose the initial azimuthal angle of the hot spot to be -90° to ensure that when the spot first passes in front of the black hole relative to our line of sight it hasn't appreciably sheared. The initial radius was chosen to be close to the ISCO since this is where a hot spot would be expected to be found, which was recently seen in [Gravity Collaboration et al. \(2018\)](#). Furthermore, an exploration of how radius and spin effects the images will be discussed below. The accretion flow parameters α and κ were chosen to be close to a perfect Keplerian motion since we anticipate this will be the motion of the accretion disk in Sgr A*. The radial size of the spot R_s was chosen to be $0.5M$ to test the ability of the EHT to resolve spots similar to the beam size of the EHT. The total observation time was $2T_K$ at the initial spot location, where T_K is the Keplerian orbital period and for a spot at $5.3r_{\text{ISCO}} = 53$ min. Each of the twelve frames has a resolution of 64×64 . While higher resolutions can be used, we made this choice for two reasons: First, it is low enough to allow for movies to make in a reasonable timescale for parameter estimation. Second, higher resolutions did not appreciably impact parameter estimation, which is presented in the next section.

³Since we assumed that hot spots are created from local microphysics, i.e. from fast magnetic reconnection, we expect the hot spot to appear suddenly and be localized initially.

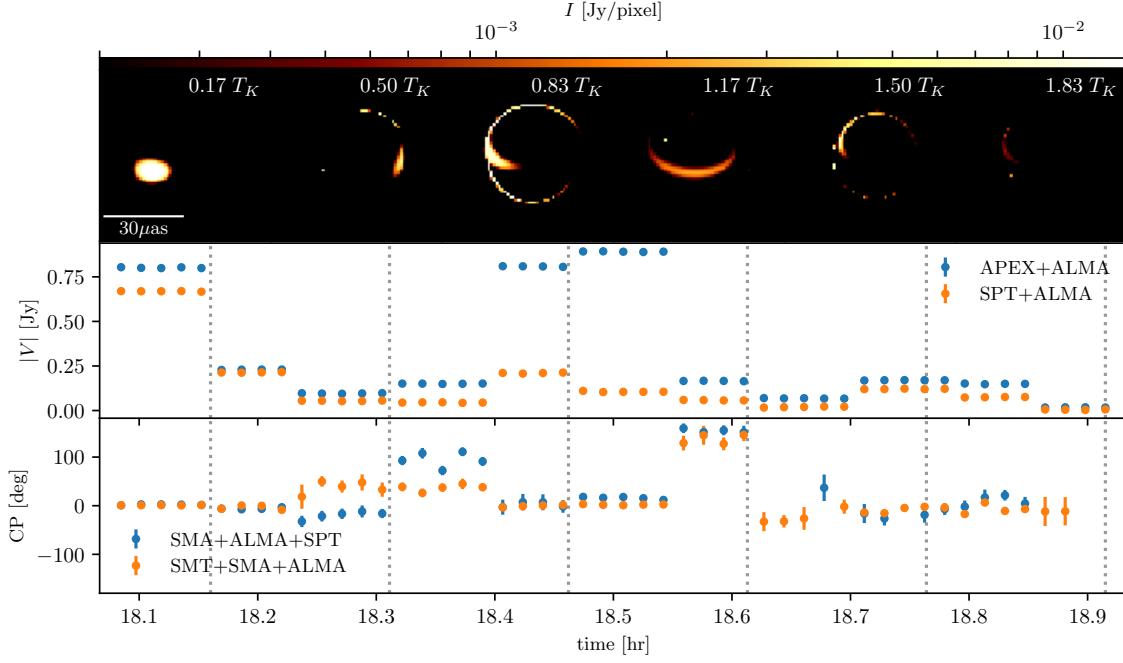


Figure 6.2: VLBI observables for the twelve frame movie shown in [Figure 6.1](#). The top panels show a subset of the 12 frames in logarithmic scale. The middle panels show synthetic VM at the APEX+ALMA (blue) and SPT+ALMA (orange) for the 2017 EHT observation baselines as a function of time. The bottom panels show the CP as a function of time at the SMA+ALMA+SPT (blue), and SMT+SMA+ALMA (orange) triangles. The gray dotted lines show where the time of movie snapshots in the top panel are taken. Note that the discrete jumps in the observations are due to the fact that we only used a 12 frame movie.

6.3 Shearing hot spots with the EHT

To assess whether the EHT has sufficient fidelity to recover a shearing hot spot's parameters, we need to first convert our ray-traced movies, such as the one shown in [Figure 6.1](#), to the VLBI data that, e.g., the EHT 2017 array will observe. The EHT, like all VLBI arrays, measure not the intensity map of an image but instead quantities associated with its Fourier transform, namely the complex visibilities. Each pair of EHT stations form a baseline for which a complex visibility is recorded. In this section, we will describe the procedure we used to convert our images into EHT synthetic data. In the first part, we describe what observables the EHT measures in more detail, and how we create synthetic

data for the movie shown in [Figure 6.1](#). In the second, we report a Bayesian MCMC parameter estimation exercise and analyze the EHT’s ability to reconstruct shearing hot spot parameters from the simulated data. In the third part, we will analyze the impact of two systematics (diffractive scattering and a background RIAF) will have on the parameters posterior distribution. In all cases, we find that the EHT can recover all the true hot spot parameters to sub-percent precision at 95% confidence about the median.

6.3.1 Creating Synthetic EHT Data

To explore how well the EHT can recover hot spots, we first need to convert the movie into interferometric EHT data. The EHT measures complex visibilities defined by,

$$V_{ij} = \int d\alpha d\beta I(\alpha, \beta) e^{2\pi i(\alpha u + \beta v)}. \quad (6.13)$$

Due to phase calibration issues from atmospheric turbulence, the phases of individual stations are practically randomized. To get around this, visibility amplitudes (VA), $|V_{ij}|$, are used, for which calibration or gain uncertainties are typically around 10% and can effectively be modeled using the same gain marginalization procedure as in [Appendix A](#). To recover some information about the complex phase of the visibilities closure phases (CP) are constructed

$$\Phi_{i,j,k} = \arg(V_{ij}V_{jk}V_{ki}), \quad (6.14)$$

which are just the sum of the phases of a triplet of visibilities. Because the baselines close, i.e., they form a triangle $(u, v)_{ij} + (u, v)_{jk} + (u, v)_{ki} = 0$, all station-specific gain errors vanish from the CPs.

In this chapter, we will use VA and CP data to explore the ability of the EHT 2017 to reconstruct shearing hot spot parameters at 230GHz. To convert spot intensity maps into interferometric data we use the *Event Horizon Telescope Imaging Library* (`eht-imaging`)⁴; ([Chael et al., 2016b, 2018b; Chael et al., 2019](#)). `eht-imaging` provides the ability to convert intensity maps into VA and CP data that uses the EHT 2017 array configuration including the correct baseline information, atmospheric thermal noise, and gain errors. This allows us to directly sample the image at the baselines the EHT will sample in a given observation window. We created an observation at 51544MJD starting at 1800hr with scan and integration time of 61s and 31s respectively. The total observation was 53 min, which corresponds to two Keplerian orbits for a spot at $5.25M$. The specific

⁴<https://github.com/achael/eht-imaging>

baselines we used are shown by the white points in [Figure 6.7](#). We also include Gaussian thermal noise in all observations, where the error, σ_{ij} , for the (i, j) baseline is determined by

$$\sigma_{ij} = \frac{1}{0.88} \sqrt{\frac{\text{SEFD}_i \text{SEFD}_j}{2 t_{\text{int}} \nu_{\text{bw}}}}, \quad (6.15)$$

where ν_{bw} is the bandwidth of the observation which we set to $4 \times 10^9 \text{ Hz}$, $2\nu_{\text{bw}}t_{\text{int}}$ is the number of independent samples of the two station baseline (i, j) , and the $1/0.88$ factor is due to two-bit quantization ([Thompson et al., 2017](#)). Furthermore, the SEFD (System Equivalent Flux Density) of each station are provided by `eht-imaging`. While gain errors can be included by `eht-imaging`, we will ignore their impact for simplicity.

When creating the synthetic VA and CPs we placed a signal to noise ratio (SNR) cut of 2 on every baseline and debiased the VA's according to

$$\mathcal{V}_{ij} = \sqrt{|V_{ij}|^2 - \sigma_{ij}^2}. \quad (6.16)$$

This allows us to approximate the error distributions of the VA and CP as a Gaussian to $< 10\%$ accuracy ([Thompson et al., 2017](#); [Broderick et al., 2020](#)). Furthermore, and SNR cut of 2 only removes a handful of measurements. The resulting visibility amplitudes and closure phases for a few baselines as a function of time are shown in [Figure 6.2](#). Short baselines, which probe large scales, show modest variations associated with the hot spot, consistent with the light curve. In contrast, long baselines, which probe small scales, exhibit large variations, associated with the rapidly varying structures within the image. In both cases, the variations are easily identifiable, substantially exceeding the thermal noise.

6.3.2 Extracting spacetime and spot parameters with Themis

To quantitatively examine the ability of the 2017 EHT array to recover and constrain hot spot parameters we use MCMC to recover the posterior distribution. To accomplish this, we used the software suite `THEMIS`. `THEMIS` is a highly extensible parameters estimation framework that was developed to deal with modeling and feature extraction of EHT observations. Furthermore, it can easily accommodate time variable structures. For more information about `THEMIS` see [Broderick et al. \(2020\)](#). For modeling, we used `THEMIS`' non-marginalized Gaussian likelihoods for both the visibility amplitudes and closure phases.

We expect EHT observations during quiescent, non-flaring periods to place strong constraints on the orientation of the black hole spin and the images azimuthal orientation. Therefore, we hold the black hole inclination, $\cos \Theta$, and image position angle, ξ , fixed

to their “true” values during each MCMC run. The reasoning behind this is that when analyzing a real data set, we expect that imaging studies will provide a prior estimate for both parameters. In principle, these parameters could be allowed to vary, however, we found that it does not alter any of the results below.

Flat priors were chosen for $a_* \in (0, 1)$, $R_s \in (0.01, 10)$, $t_0 \in (-200, 200)$, $r_0 \in (1.5M, 20M)$, $\phi_0 \in (-\pi, \pi)$, $\alpha \in (0, 1)$, $\kappa \in (0, 1)$, and n_0 used a logarithmic prior, ranging from 10^3 to 10^{12} . For the MCMC sampler we used the parallel tempered affine-invariant sampler originally detailed in Goodman & Weare (2010); Vousden et al. (2016), with 48 walkers, 6 tempering levels, with temperature swaps every 50 MCMC steps. To speed up convergence, the walkers were started at the true values of the model. A single run for 1000 MCMC steps took 150 000 core hours, on the Calcul Quebec and Compute Canada cluster Mp2.

The joint parameter posterior probability distributions are shown in Figure 6.3. Every spot parameter is recovered with sub-percent accuracy. For example, the median spin and its 95% confidence interval was $a_* = 0.50002^{+0.00017}_{-0.00023}$. Every marginalized distribution was single-peaked, showing no apparent degeneracies using the EHT 2017 array. The minimum reduced chi-square was found to be 1.0001 with 1640 degrees of freedom. Therefore, we conclude that the EHT 2017 array can accurately recover isolated shearing hot spots around Kerr black holes with high precision and accuracy. While these results are encouraging, we have ignored all potential systematics except the thermal noise present in the array. In the next section, we will study the impact of a few potential systematics that are important for Sgr A*.

6.3.3 Adding Systematics – Scattering and Background flows

In practice, the idealized observations described in the previous section are not directly applicable to EHT data, which are subject to a variety of additional systematic effects. Therefore here, we analyze how some of these systematics modify our conclusions. We focus on two systematics that are expected to dominate the error budget: diffractive scattering and a background accretion flow. Additionally, we analyze the impact of different inclination angles of the accretion disk on our parameter estimation.

Diffractive Scattering

Emission from SgrA* is scattered by interstellar electrons (Bower et al., 2006; Johnson et al., 2018; Issaoun et al., 2019). This both blurs the image (diffractive scattering) and

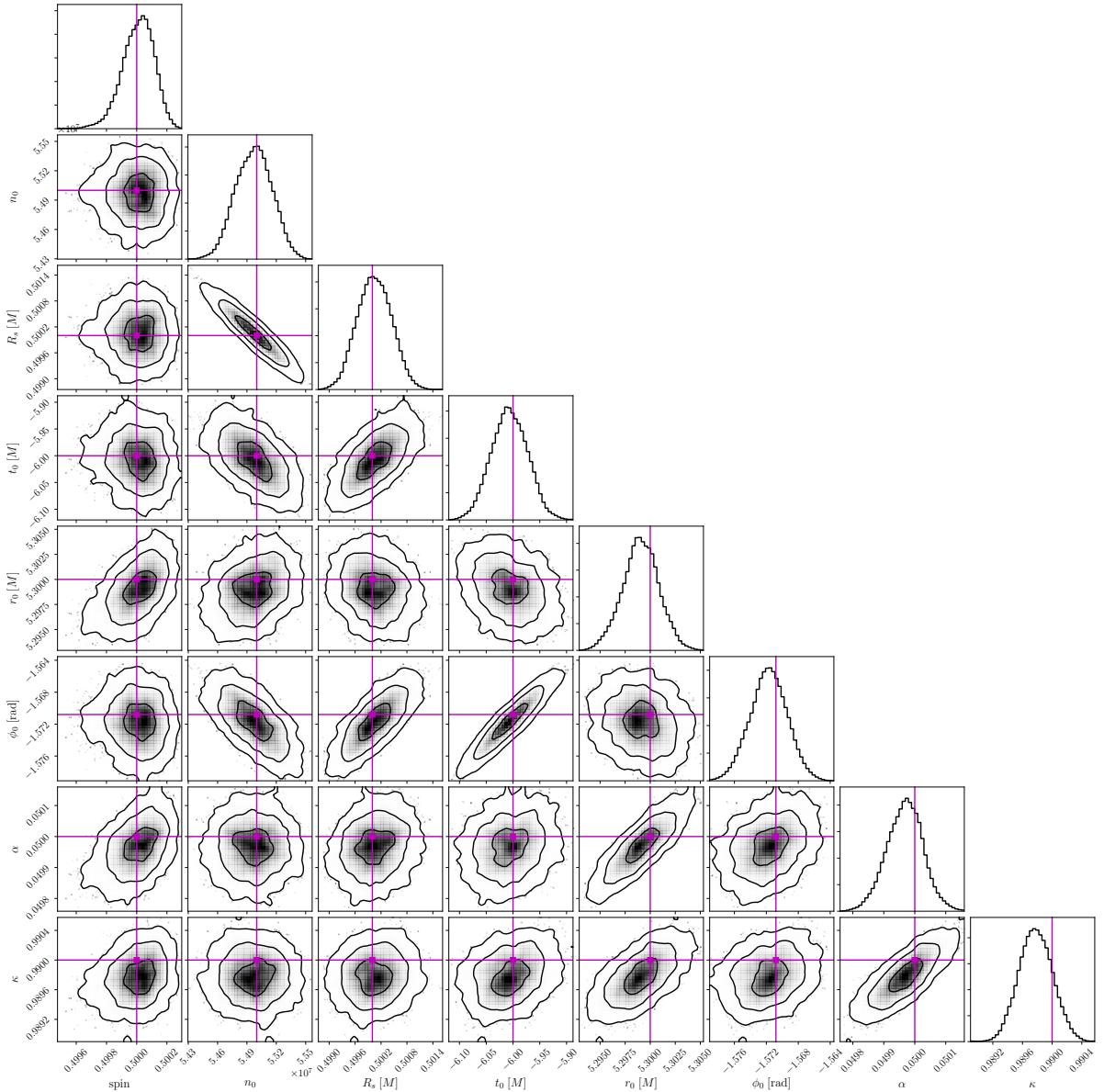


Figure 6.3: Joint posterior probability distribution of shearing spot. The purple lines and points show the true values of the model, which is shown in [Figure 6.2](#).

stochastically lenses the image (refractive scattering). We consider the implications of the

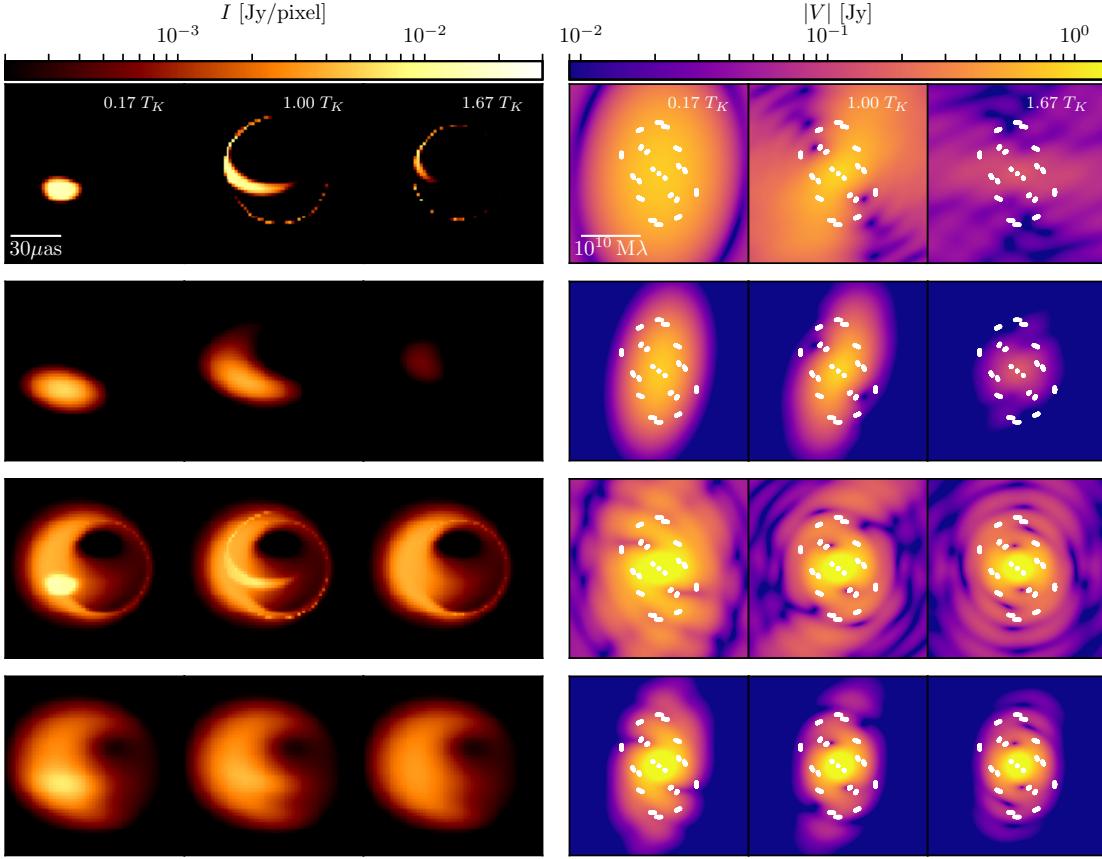


Figure 6.4: Comparison of movies snapshots (left) and their corresponding visibility amplitudes (right) as different systematics are added. The white dots denote the (u,v) points sampled using the EHT array configuration described in the main text. The top figures are the base case with no systematics. The second from the top is with diffractive scattering, second from the bottom is with a RIAF and no scattering, and the bottom is with both a RIAF and scattering.

former for the reconstruction of shearing hot spots here, leaving the latter for future work. Diffractive scattering has the effect of washing out any structure below the scale of the blurring kernel. For our kernel, we use the empirically determined, wavelength-dependent

asymmetric Gaussian kernel from [Bower et al. \(2006\)](#) with parameters,

$$\begin{aligned}\theta_{\text{maj}} &= 1.309 \left(\frac{\lambda}{1 \text{ cm}} \right)^2 \text{ mas}, \\ \theta_{\text{min}} &= 0.64 \left(\frac{\lambda}{1 \text{ cm}} \right)^2 \text{ mas},\end{aligned}\tag{6.17}$$

where $\theta_{\text{maj,min}}$ is the FWHM of the semi-major/minor axis of the Gaussian scattering ellipse and $\psi = 78^\circ$ its orientation. At 230GHz this corresponds to a semi-major axis FWHM of $22\mu\text{as}$. The impact of diffractive scattering on the image is shown in the left panels of [Figure 6.4](#). As expected, the diffractive scattering removes structure smaller than the typical kernel size of the image. For visibility amplitudes (right panels of [Figure 6.4](#)), this corresponds to the suppression of visibility amplitudes at high baselines length. The impact on the direct EHT observables for the hot spot movie in [Figure 6.1](#) is shown in [Figure 6.5](#). Here we see that at long baselines, i.e. the SPT+ALMA baseline, variations in the VM are damped. Note that since the scattering Kernel is a Gaussian the closure phases are not modified.

RIAF Background

While hot spots can contribute substantially to the image flux, the main source of emission, the accretion disk, will typically dominate. Where it does not, its opacity will still obscure the hot spot emission. To include the impact of an accretion disk we included a radiatively inefficient accretion flow (RIAF) model from [Broderick et al. \(2016\)](#) fitted to past proto-EHT observations of Sgr A*. [Figure 6.4](#), demonstrates how the RIAF background impacts hot spot movies. The impact is twofold. One, we see that regions, where the emission is very dim, is washed out by the background RIAF. Secondly, and most importantly, is the impact of optical depth from the background accretion flow. This effect is especially pronounced in the Doppler boosted region of the disk. In this region, the spot becomes entirely washed out after it passes through the ISCO. This suggests that hot spots appearing inside the ISCO will be much harder to observe with the EHT.

In terms of EHT observables, we see how the RIAF impacts the visibilities and images in Figures [6.4](#) and [6.5](#)). After adding the RIAF, the spot brightness above the background drops from 0.8Jy to 0.5Jy from optical depth. The impact of optical depth is even more pronounced after the spot makes two complete orbits⁵, as is seen in the third panel of

⁵Since the spot has radial motion it completes more than two orbits during the observation time.

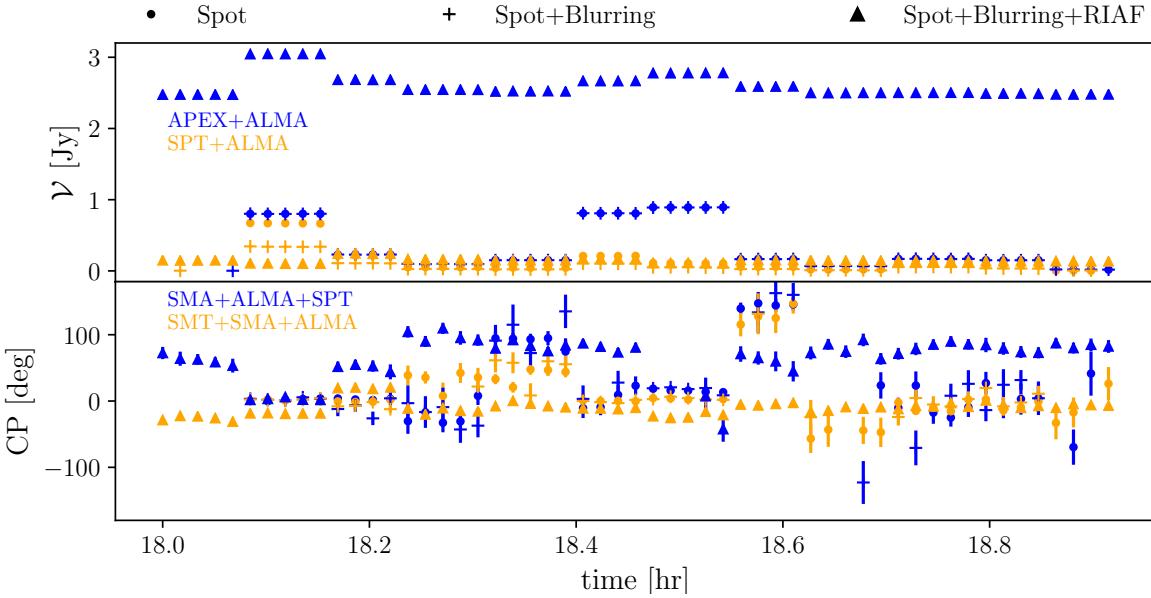


Figure 6.5: Impact of scattering and a background RIAF on the VM (top) and CP (bottom) as the spot falls into the black hole. We again use the same 12 frame movie parameters as shown in [Figure 6.2](#)

the third column of [Figure 6.4](#). Before adding the RIAF, the hot spot extends across the entire face of the black hole and the secondary emission is visible. After, the hot spot is practically invisible.

Another way to see this impact is by analyzing how the light curve of the spot changes as the initial radius of the spot moves inwards, which is shown in [Figure 6.6](#). To assure a fair comparison as the initial radius of the hot spot is changed, we decrease the density constant n_0 too. This ensures the maximum brightness of the spot fixed to $\sim 0.5\text{Jy}$. Analyzing [Figure 6.6](#) as the starting radius of the hot spot is moved inwards, it moves into the optically thick region sooner. This leads to the second orbit of the spot is increasingly obscured. As we will see below, for hot spots starting inside the ISCO, this negatively impacts the ability of the EHT to recover hot spots.

Impact of systematics on parameter estimation

To estimate the impact of the systematics on the parameter estimation, we use the procedure described in [Section 6.3.2](#), with identical starting parameters, priors, and sampler

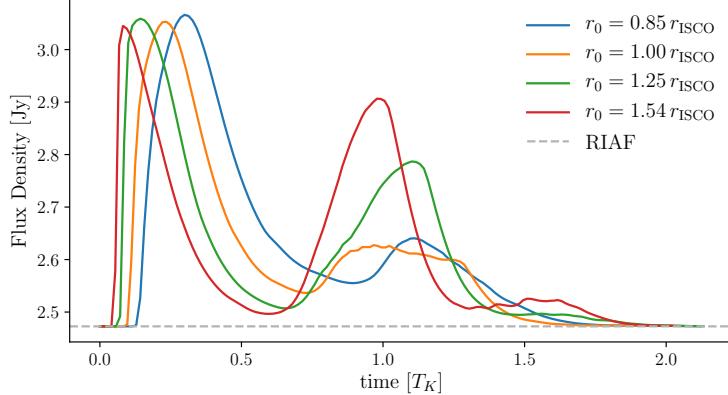


Figure 6.6: VLBI lightcurves for a shearing spot with blurring and a RIAF background, around a black hole with spin parameter $a_* = 0.5$, at the four different radii specified in Table 6.1. The units of the x-axis are given as a fraction for the Keplerian orbital period, T_K , at the initial radius for the respective spot.

options, but including the RIAF and blurring to the model when applicable. The impact of the systematics on the posterior distribution is shown in Figure 6.7. Blurring does not appear to impact the posteriors substantially. One reason for this is that diffractive scattering doesn't change closure phases. Additionally, looking at Figure 6.4, blurring is a multiplication of the VA and can easily be inverted through modeling, since the kernel has no nulls in visibility space. However, when the background RIAF and diffractive scattering are both included the posteriors do broaden. For the black hole spin, the range of the inferred values increases by roughly a factor of two. However, we still have sub-percent precision, finding that $a_* = 0.5001^{+0.00073}_{-0.00129}$. Therefore, even with blurring and a RIAF background, the EHT 2017 array can recover hot spots and sub-horizon-scale physics to high accuracy.

In the absence of a background RIAF, all spacetime constraints arise from the shearing hot spot. In the presence of a RIAF, the morphology of the quiescent accretion flow provides additional information (Johannsen et al., 2016). Thus, we seek to assess the improvement in the spin measurement, arising from the inclusion of the hot spot relative to the background RIAF. Namely, if the spots are improving the measurement of spin, we would expect the bound on the spin to improve relative to just fitting a RIAF background during the same observation. To test this, we used the same array configuration as for the hot spot simulations, and then created simulated data of a RIAF with the same parameters used previously. The results are shown by the black curves in Figure 6.7, which compares

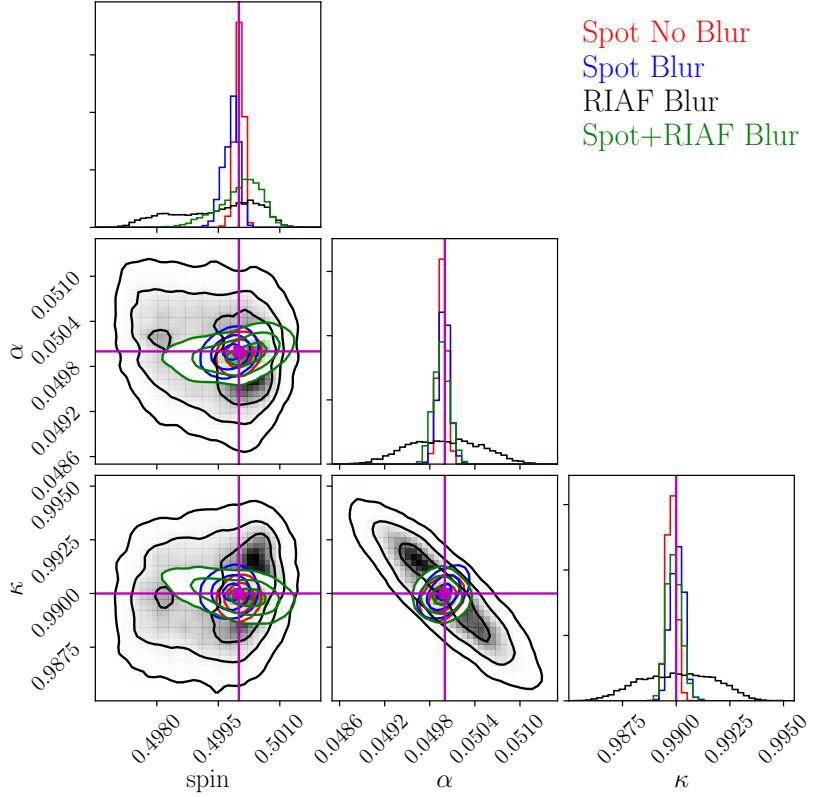


Figure 6.7: Comparison of the joint-probability distribution of a shearing hot spot with different systematics. The black contours show represent the base model without diffractive scattering and a background RIAF. Blue the same spot but with diffractive scattering, and red if the same spot with diffractive scattering and a RIAF background. The model parameters are: spin of 0.5 with a viewing inclination of 60° , accretion flow parameters are set of $\alpha = 0.05$ and $\kappa = 0.99$ (near-Keplerian), and the spot was initially placed at $1.25 \cdot r_{\text{ISCO}} = 5.3 M$ with a azimuthal angle of -90° and $\xi = 0^\circ$.

the joint probability distribution for the spin and two accretion flow parameters. For the case of the RIAF, we find the spin is given by $a_* = 0.49964^{+0.00128}_{-0.00213}$. The 95% errors are then 70% larger than the hot spot and RIAF model. Therefore, we indeed see that catching a flare does improve EHT measurements of spin.

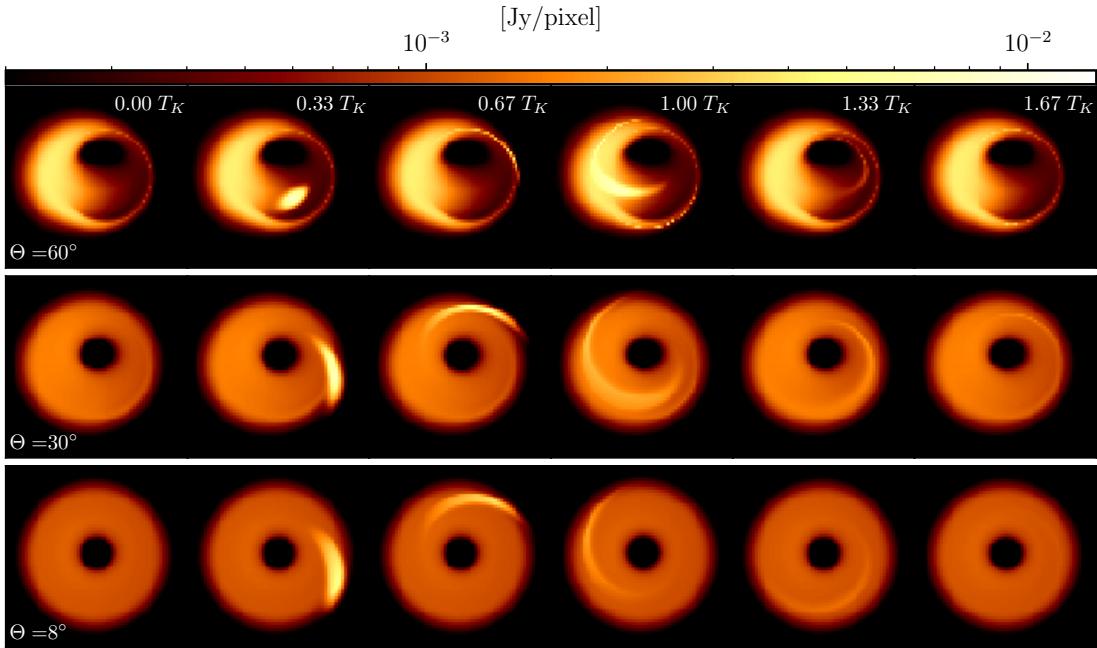


Figure 6.8: Comparison of hot spot motion embedded in the [Broderick et al. \(2016\)](#) best fit RIAF model, over different accretion disk inclination angles $\Theta = 60^\circ, 30^\circ, 8^\circ$. As the inclination angle becomes smaller lensing is suppressed since optical depth from the accretion disk becomes large.

6.3.4 The impact of disk inclination

We have shown that the inclusion of a scattering screen and background accretion flow does not drastically alter our ability to extract hot spots. However, most of the conclusions so far have assumed that the inclination of the accretion disk is $\Theta = 60^\circ$. While this angle does match what [Broderick et al. \(2016\)](#) found for Sgr A*, the uncertainty in the inclination is quite large. Furthermore, [Gravity Collaboration et al. \(2018\)](#) found that the inclination angle of the orbital plane of the hot spot motion was $\sim 30^\circ$ during a flare. [Figure 6.8](#) illustrates how the inclination angle changes the morphology of an image. As the inclination angle decreases, the impact of the lensed emission from the hot spot is suppressed since the disk becomes optically thick. Additionally, the variability of the light curve becomes subdued since the hot spot doesn't “disappear” behind the black hole at $\Theta = 30^\circ, 8^\circ$.

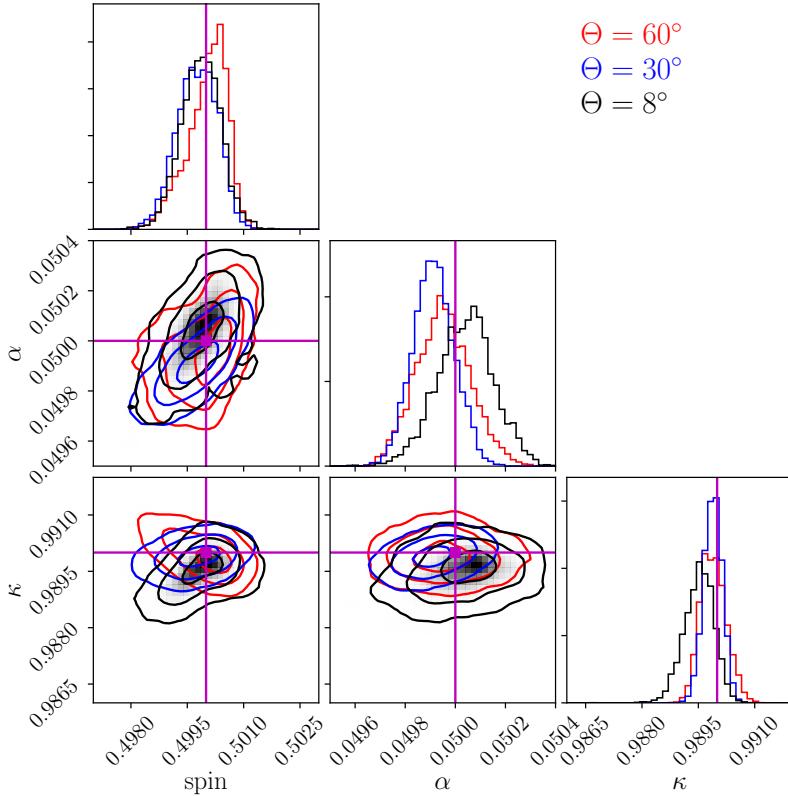


Figure 6.9: Comparison of the joint-probability distribution of shearing hot spots with a background RIAF and diffractive scattering at different inclination angles. *Red* represents the standard inclination angle used in this chapter of $\Theta = 60^\circ$ and is the same posterior as the green curve in Figure 6.7. *Blue* is for $\Theta = 30^\circ$ and is roughly the inclination found by Gravity Collaboration et al. (2018). *Black* is for $\Theta = 8^\circ$. In all instances the measurement of spin is similar to the standard $\Theta = 60^\circ$ case.

To analyze how disk inclination impacts hot spot measurements, we again followed the same procedure as above. That is, we created a twelve frame movie with scattering and background RIAF, and used `eht-imaging` to create a simulated dataset with the same array configuration as the previous experiments. The results of the parameter estimation are shown in Figure 6.9. Figure 6.9 demonstrates that the inclination angle has a negligible effect on the ability of the EHT to extract hot spots when compared to the $\Theta = 60^\circ$, again recovering spin to sub-percent precision.

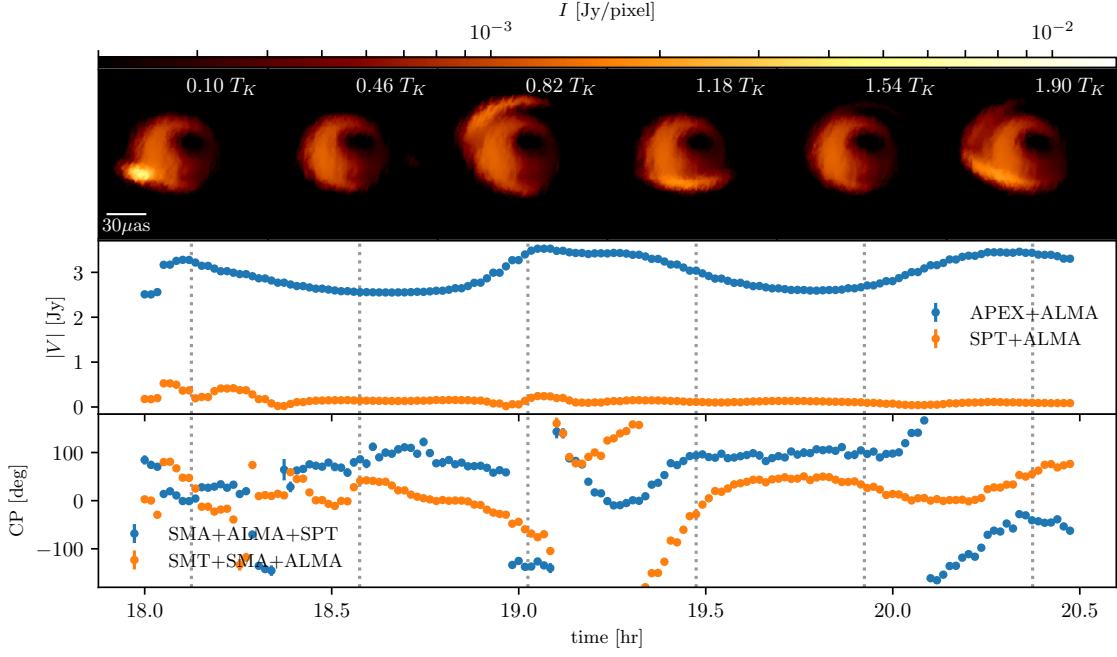


Figure 6.10: Example of a spot movie with refractive scattering and a RIAF background included. The movies uses 100 frames over a 2.5hr window. The model parameters are: $a_* = 0$, $\cos \Theta = 0.5$, $R_s = 0.5$, $n_e = 5.5 \times 10^7$, $t_0 = -6M$, $r_0 = 10M$, $\phi_0 = -90^\circ$, $\alpha = 0.05$, $\kappa = 0.99$, $\xi = 0^\circ$. The top panel shows a selection of frames of the movie the times specified by the grey dotted lines. The middle and bottom panel show the VA and CP at a couple of baselines and triangles.

Future systematics to consider

We have shown that the systematics included in this chapter does not seem to impact the ability of the EHT to recover hot spot parameters. However, other systematics need to be considered in future work. These include gain errors, refractive scattering, and variable background effects.

Even after array calibration, it is expected that there will be residual 10% gain errors in observations ([Event Horizon Telescope Collaboration et al., 2019c; Broderick et al., 2020](#)). In [Broderick et al. \(2020\)](#), a gain mitigation technique was developed that was able to marginalize the impact of gains on parameter estimation and was applied in [Event Horizon Telescope Collaboration et al. \(2019f\)](#) and typically increased posterior width by a factor of a few. Extrapolating from [Event Horizon Telescope Collaboration et al. \(2019f\)](#), we do

Table 6.1: Spacetime & Hot Spot Parameters

spin 0			spin 0.5			spin 0.9		
r_0/r_{ISCO}	$r_0 [M]$	Time [hr]	r_0/r_{ISCO}	$r_0 [M]$	Time [hr]	r_0/r_{ISCO}	$r_0 [M]$	Time [hr]
0.85	5.00	0.83	0.85	3.60	0.54	0.85	1.98	0.27
1.0	6.00	1.1	1.0	4.23	0.68	1.0	2.33	0.33
1.25	7.50	1.5	1.25	5.30	0.94	1.3	3.00	0.45
1.5	9.00	2.0	1.54	6.50	1.2	1.7	4.00	0.66

not expect that gains will then provide a significant obstacle to hot spot reconstruction.

While we have included diffractive scattering in this chapter, Sgr A* is also refractively scattered (Bower et al., 2006; Johnson et al., 2018). Refractive scattering effectively adds small scale structure to the movie impacting long-baseline visibilities. However, we don't expect this to form a barrier to hot spot reconstruction for two reasons. One, the timescale of the spot evolution is much shorter than the dynamical time scale of the scattering screen. The scattering timescale set by the orbital motion of the earth around the galactic center and is over hours, while hot spot changes on the order of minutes. Therefore, we can effectively treat the scattering screen as static during a flare. Second, as Figure 6.10 demonstrates, the scale of the scattering scintillation is typically on much smaller scales than hot spots. This is due to diffractive scattering, which smears the hot spot to scales much larger than the refractive scintillation. Taken together, this suggests that while scattering mitigation is important, it should not significantly alter the results presented in this section.

Sgr A* displays consistent small-scale variability (Witzel et al., 2018), which is presumed to arise from turbulence and shocks in the accretion disk. General relativistic magneto-hydrodynamics (GRMHD) simulations suggest that we consistently expect small-scale fluctuations in the accretion disk. For bright flares, this becomes less significant as a single region presumably dominates the emission. Nevertheless, to model the impact of this, we could include numerous sub-dominant spots, to model GRMHD turbulence, and then only attempt to recover the bright flare. Additionally, we could inject our hot spot model into a GRMHD simulation and then attempt to recover it⁶.

⁶GRMHD simulation struggle to produce these flares since they typically ignore the microphysics and plasma resistivity needed to produce fast reconnection events.

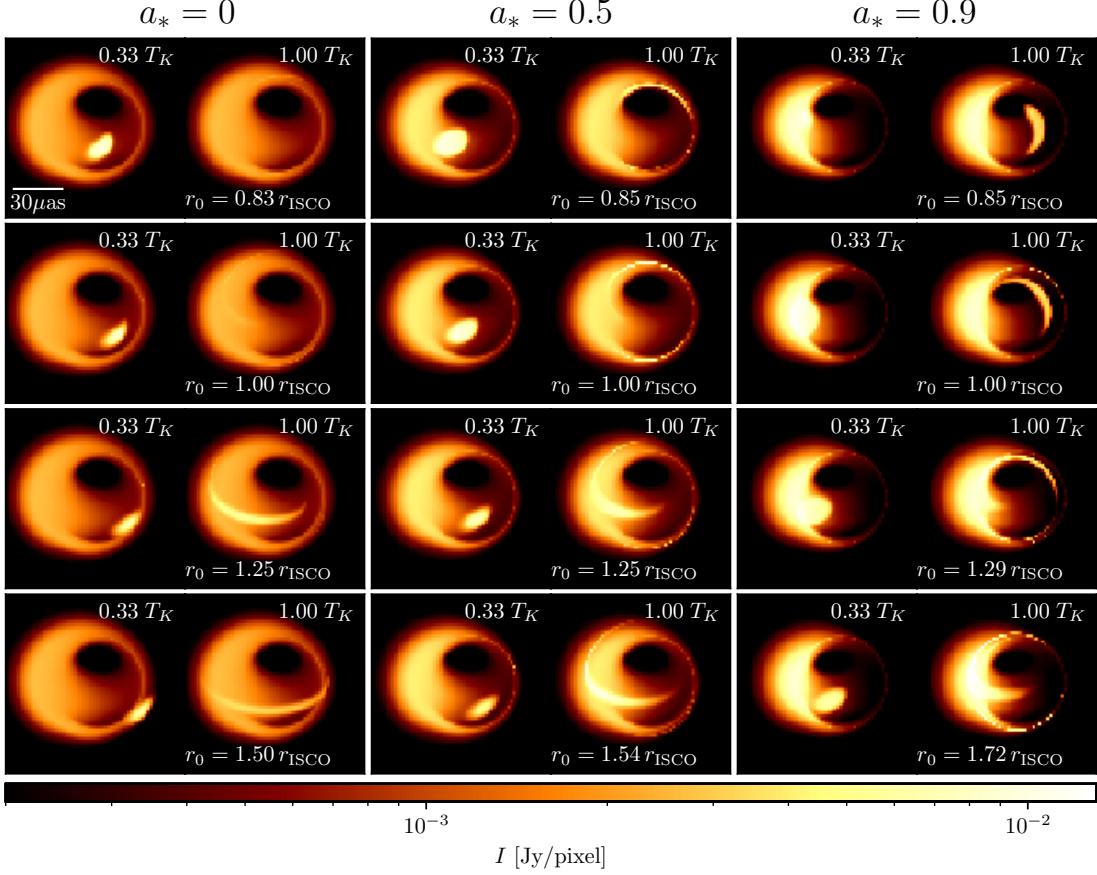


Figure 6.11: Impact of changing spin parameter (increasing from left to right) and initial radius (increasing from top to bottom) of a shearing hot spot with a RIAF background, with colors in log-scale, showing intensity per pixel.

6.4 Spacetime tomography

The frequency of flaring states in Sgr A* depends on the wavelength of the observations. At NIR, Sgr A* has a significant flare ~ 4 times per day (Genzel et al., 2003; Eckart et al., 2006; Meyer et al., 2009, 2014; Hora et al., 2014), while only a quarter of those typically have an X-ray counter part (Baganoff et al., 2001; Eckart et al., 2004; Marrone et al., 2008; Porquet et al., 2008; Do et al., 2009; Neilsen et al., 2013; Mossoux et al., 2015). Sub-mm occur 1–4 times a day (Marrone et al., 2008; Dexter et al., 2014). For the EHT however, it is not entirely clear whether the NIR/X-ray or sub-mm rate is relevant, given that the

observations are at horizon scales. Either way, for any of the flaring rates of Sgr A*, the EHT will likely capture at least one flare per observational cycle. This implies that the EHT will measure multiple flares in the next few years.

These flares permit the opportunity to reconstruct the spacetime parameters in a position-dependent fashion, e.g., map the spacetime as a function of the initial hot spot radius. The bundle of light rays (i.e., null geodesics) connecting the primary and higher-order images of a given hot spot will pass through different regions of the underlying spacetime for spots launched at different orbital radii. Thus, the black hole mass and spin measurements from subsequent flaring epochs provide a spatially resolved probe of the black hole spacetime. Such a spatially-resolved spacetime probe, or tomographical map of spacetime, provides a natural test of the no-hair theorem. We will explore the limits that can be placed in practice on parameterized deviations from GR in a future publication. One caveat to note is that in this chapter we have chosen a typical model for these flares where the initial spot is in the disk. It is possible that the hot spot could form out of the plain of the disk or have significantly different accretion dynamics from those assumed in this chapter. While this does mean our tomographical map of spacetime is model dependent, it provides an additional avenue to probe spacetime on event horizon scales.

6.4.1 Constructing a synthetic tomographical map of spacetime

To explore the ability of EHT to perform spacetime tomography, we placed a series of hot spots in a Kerr spacetime varying both the initial radius of the hot spot and the spin of the black hole. [Table 6.1](#), lists the radii and spins that were considered. As the initial radius, r_0 , changes, the orbital period varies as well. To ensure that each experiment contains the same hot spot evolution, we restrict all movies to be $2T_K(a_*, r_0)$ each using 12 frames. Additionally, due to optical depth effects, the brightness of a hot spot will change as the spin and initial radius varies. Therefore, for each movie, we picked the hot spot density normalization, n_0 , such that the brightness was ~ 0.5 Jy's above the quiescent emission. This ensures our results aren't due to the brightness of the hot spot. The other parameters, Θ , R_s , t_0 , ϕ_0 , α , κ , ξ , were held fixed between experiments and set to the same values in [Section 6.3.2](#). Finally, each movie includes a background RIAF and diffractive scattering.

[Figure 6.11](#), shows how the intensity maps of each movie in [Table 6.1](#), with the considerations in the above paragraph. The same subset of frames, in terms of T_K , is chosen for each movie. As the initial radius of the spot and spin of the black hole change so does that image by significant amounts. Furthermore, for small r_0 it becomes difficult to see the hot spot after one orbit due to the optical depth of the accretion disk. As we will see below, this can impact the ability of the EHT to recover hot spots close to the black hole.

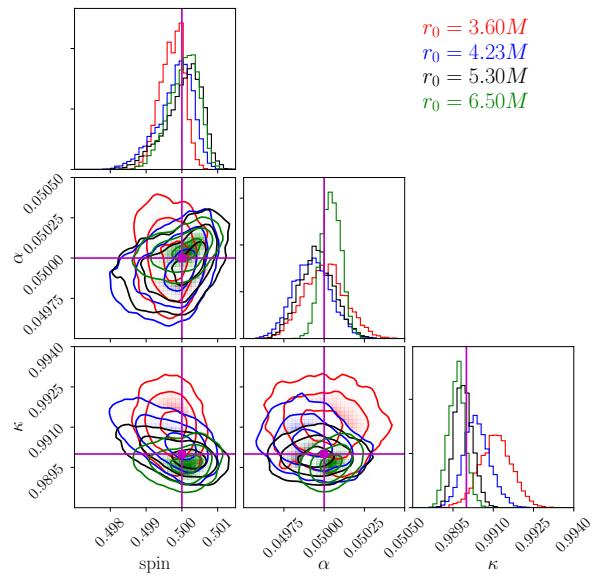
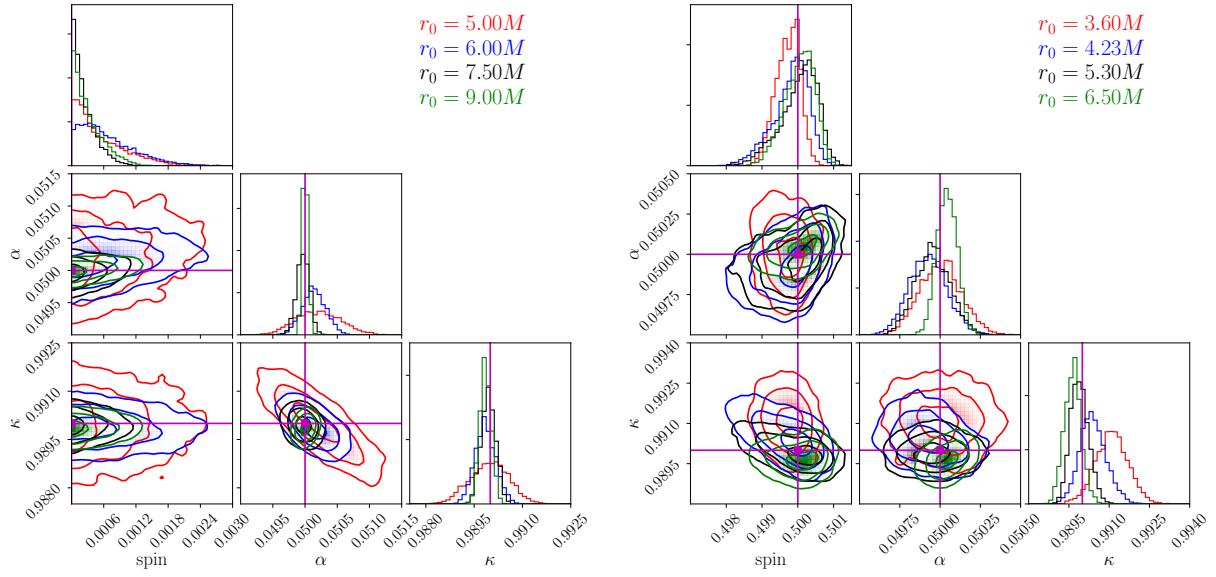


Figure 6.12: Joint probability distributions of the spin a_* , and two accretion flow parameters α, κ , for $a_* = 0$ (top left), $a_* = 0.5$ (top right) and $a_* = 0.9$ (bottom).

To create the synthetic EHT data, we used the same procedure described in Section 6.3.1. Namely, we used the same scan and integration time, observation frequency and bandwidth, and the same MJD and start time of the observation. Due to the movies being different lengths, each observation will have a different number of data points. This is a realistic simulation of actual spot observations since the duration of a flare sets the time interval we are interested in modeling.

6.4.2 Results

For parameter estimation, we used the same procedure described in Section 6.3.2 and 6.3.3. The joint probability distributions for each run are shown in Figure 6.12 for the spin 0, 0.5, 0.9 cases. Every experiment was able to recover the spot parameters to sub-percent levels using 95% confidence levels about the median of the marginalized posteriors. Furthermore, the peaks of the joint-probability distributions are all statistically consistent with the true values of the model, which is shown by the purple line.

Contrary to naive expectations based on spacetime considerations alone, the estimation of the black hole spin does not substantially improve as the spot moves closer to the black hole. This is because the optical depth from the accretion disk suppresses the intensity of the spot dramatically as it falls past the ISCO. For spots farther out, more of the hot spot is visible, giving much better constraints on the black hole spin.

Fixing the spin, we can associate each flare with a characteristic radius, i.e., the initial radius⁷. The results of each column in Table 6.1 forms a tomographical map of the given spacetime and is shown in Figure 6.13. General relativity predicts that for a given spin, each flare must lie on a horizontal line in Figure 6.13. If there was evidence of curvature, either the astrophysical model, e.g., accretion flow dynamics, is incorrect or that nature may deviate from GR near horizon scales.

In summary, we have found that the EHT can tomographically map spacetime and accretion flow dynamics near the event horizon, providing a new test of GR in the strong gravity regime. While this test is not truly model-independent, it does provide an additional avenue to test the no-hair theorem, independent of others such as the black hole shadow size (Johannsen et al., 2016; Psaltis et al., 2015, 2016).

⁷While other choices are possible, the initial radius is the simplest and is a model parameter

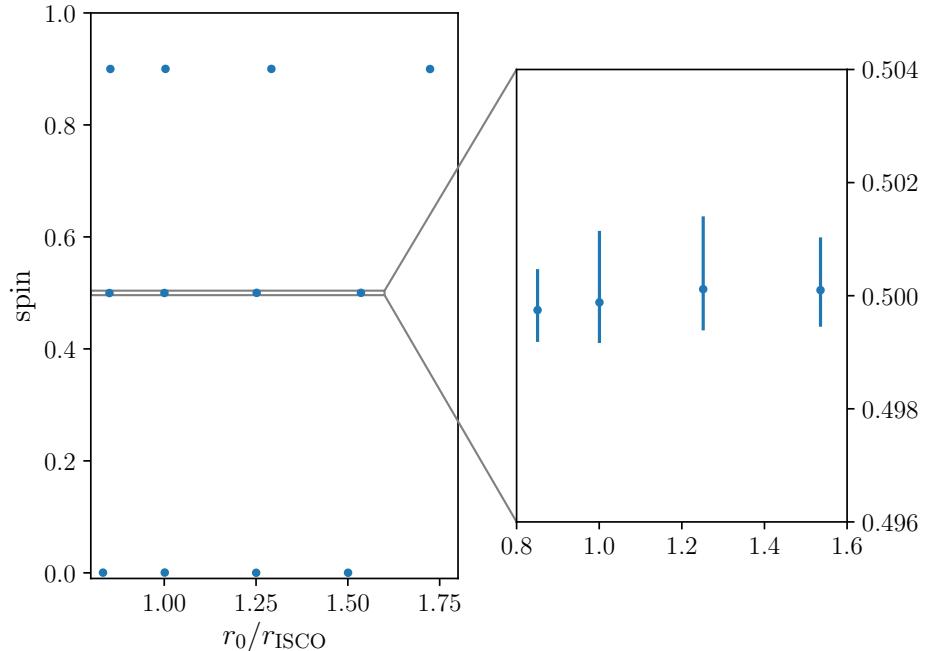


Figure 6.13: Plots of recovered spin and initial radii parameters of shearing spots, with the effects of diffractive scattering and a background RIAF included.

6.5 Conclusions

Resolving structural variability on timescales of minutes to hours presents an opportunity to probe accretion processes and gravity on horizon scales. This is especially true for Sgr A* that displays dramatic flaring events every 1–3 days. Gravity Collaboration et al. (2018) associated these flares with hot spots in the accretion disk thought to have arisen from magnetic reconnection events in the accretion disk and predicted over a decade ago Broderick & Loeb (2005, 2006). It is expected that these spots will expand and shear as they traverse around the black hole since they are embedded in an accretion flow. Therefore, we introduced a novel semi-analytical shearing hot spot model enabling us to perform parameter estimation studies with the EHT.

Using said model, we have shown that the 2017 EHT array can recover the hot spot parameters, such as spin, to sub-percent precision. Without any systematics, we were able to recover the spin to 0.1%. Including diffractive scattering and a background accretion

flow (see [Figure 6.7](#)), we can recover spin to 0.4%, and were able to show that this results did not depend on the black hole inclination. Furthermore, we were able to recover the spin to 0.05%–0.5% for a variety of different initial radii, as can be seen in [Figure 6.12](#). By combining each of these results, we have demonstrated how observing hot spots naturally leads to a notion of mapping out the radial structure of accretion and spacetime at horizon scales and forming a tomographical map of spacetime. Furthermore, these results demonstrate the power of variability as a tool to test GR. In [Chapter 5](#), minimal constraints on the nature of the black hole were made, other than it was consistent with expectations from GR. However, no leverage on the spin or non-GR effects could be inferred.

In future works, we plan to analyze how additional systematics impact the results in this chapter. Additionally, hot spots are not the only source of variability in Sgr A*. Turbulence and shocks in the accretion disk are thought to be responsible for most of the small-scale variability seen in Sgr A*. Therefore, we plan on analyzing the impact of such variability on hot spot parameter estimation. This explores whether multiple hot spots could be used to model the turbulence and shocks in GRMHD simulations and their impact on the results in this chapter.

Chapter 7

Conclusion

Understanding gravity and matter on horizon scales is a critical open question in physics. With the advent of the Event Horizon Telescope, we are now entering an era where probing this question is possible. In the first EHT publications, the mass and the polarization structure of M 87 were measured. In this thesis, we have demonstrated the current and future potential of the EHT to measure the nature of accretion and gravity on horizon scales.

One of the key systematics in understanding gravity on horizon scales is characterizing uncertainty. The EHT is a very sparse interferometer. As a result, image reconstructions are highly uncertain. [Chapter 3](#) demonstrates that the traditional imaging techniques used by the EHT and radio astronomy are unable to reliably measure image ellipticity. To account for this bias, complicated and needlessly conservative calibration procedures are required. The origin of this problem is fundamental to the methods. Namely, they are intrinsically unable to measure uncertainty.

In [Chapter 5](#), we demonstrated that Bayesian techniques do not suffer from the same problems as traditional imaging methods. Furthermore, since no calibration is required, we improved M 87’s ellipticity constraint by a factor of 3. Additionally, we found that the observed ellipticity of M 87 was consistent with the expected ellipticity of GRMHD simulations. This analysis was made possible due to the recent development of THEMIS ([Broderick et al., 2020](#)) and new sampling techniques detailed in [Chapter 4](#).

Bayesian and sampling techniques from [Broderick et al. \(2020\)](#) and [Chapter 4](#) have already been applied to other imaging problems. For instance, in [Event Horizon Telescope Collaboration et al. \(2021a\)](#) polarized Bayesian images were constructed, constraining the polarization of the on-sky image. This information was then used to constrain the state of

accretion around M 87 (Event Horizon Telescope Collaboration et al., 2021b). Currently, efforts are underway to apply the Bayesian imaging pipeline to other AGN sources in the EHT.

In the future, improving the scaling of Bayesian imaging will be essential, specifically for the next-generation EHT (ngEHT). The ngEHT could have double the number of telescopes compared to the EHT. This would quadruple the amount of data. The amount of data could be further increased if additional frequencies are included, e.g., 86 and 345 GHz. The effect of the increase in data is twofold. First, likelihood evaluation scales linearly with the amount of data, becoming a significant bottleneck. Second, as the amount of data increases, so does model complexity. In the context of Bayesian imaging, this will require a larger number of raster control points, further increasing the computational cost. To solve this problem will require new modeling and sampling tools, such as variational inference (Blei et al., 2017; Sun & Bouman, 2020; Arras et al., 2021), or sub-sampling MCMC schemes (Cornish et al., 2019).

Chapters 2–5 focused on modeling the on-sky image. However, describing the physical processes on horizon scales that generate the on-sky image provides the greatest discovery potential. For example, directly modeling the physical accretion disk would include gravitational parameters, such as mass and spin. Therefore, using these physical models would allow for direct inferences on spacetime structure (Johannsen et al., 2016). This approach has already been applied to proto-EHT observations, and as a result constrained the orientation and emission properties of Sgr A* (Broderick et al., 2016).

In Chapter 6, we demonstrated the power of modeling a variable source on horizon scales. Using a hotspot model for Sgr A* flares, we demonstrated that a single observation of Sgr A* during a flare could result in high-precision spin measurements. Furthermore, we showed how a tomographic map of spacetime near the event horizon could be constructed by observing multiple flares. Constructing such maps will be possible soon. During the 2017 EHT observation, Sgr A* was observed during an X-ray flare. Given that Sgr A* flares once per day, we expect to have observed multiple flares within a decade.

In the future, we will extend the hotspot model from Chapter 6 to infrared and X-ray flare observations (Genzel et al., 2003; Neilsen et al., 2013). While infrared and X-ray flares are known to be correlated (Ponti et al., 2017), after more than a decade of research, the relation between flares at different wavelengths is still poorly understood (Fazio et al., 2018). Thus, extending the hotspot model in Chapter 6 across multiple frequencies could help elucidate the different emission properties. Furthermore, reconstructing these events at different wavelengths would effectively form a tomographic map of *phase space* around black holes, probing the nature of black holes and AGN on horizon scales.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102, doi: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102)
- Abedi, J., Dykaar, H., & Afshordi, N. 2017, Phys. Rev. D, 96, 082004, doi: [10.1103/PhysRevD.96.082004](https://doi.org/10.1103/PhysRevD.96.082004)
- Aherne, F. J., Thacker, N. A., & Rockett, P. I. 1998, Kybernetika, 34, [363]. <http://eudml.org/doc/33362>
- Akiyama, K., Kuramochi, K., Ikeda, S., et al. 2017a, ApJ, 838, 1, doi: [10.3847/1538-4357/aa6305](https://doi.org/10.3847/1538-4357/aa6305)
- Akiyama, K., Ikeda, S., Pleau, M., et al. 2017b, AJ, 153, 159, doi: [10.3847/1538-3881/aa6302](https://doi.org/10.3847/1538-3881/aa6302)
- Almheiri, A., Marolf, D., Polchinski, J., & Sully, J. 2013, Journal of High Energy Physics, 2013, 62, doi: [10.1007/JHEP02\(2013\)062](https://doi.org/10.1007/JHEP02(2013)062)
- Arras, P., Frank, P., Haim, P., et al. 2021, M87* in space, time, and frequency. <https://arxiv.org/abs/2002.05218>
- Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27, doi: [10.3847/1538-4365/ab06fc](https://doi.org/10.3847/1538-4365/ab06fc)
- Baganoff, F. K., Bautz, M. W., Brandt, W. N., et al. 2001, Nature, 413, 45, doi: [10.1038/35092510](https://doi.org/10.1038/35092510)
- Bergmann, P. G., & Brunings, J. H. 1949, Reviews of Modern Physics, 21, 480
- Bergmann, P. G., & Goldberg, I. 1955, Physical Review, 98, 531

- Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., & Stuart, A. 2013, Bernoulli, 19, 1501 , doi: [10.3150/12-BEJ414](https://doi.org/10.3150/12-BEJ414)
- Betancourt, M. 2017, arXiv e-prints, arXiv:1701.02434. <https://arxiv.org/abs/1701.02434>
- Betancourt, M. J. 2015, Adiabatic Monte Carlo. <https://arxiv.org/abs/1405.3489>
- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAM review, 59, 65. <https://doi.org/10.1137/141000671>
- Bhattacharyya, A. 1943, Bull. Calcutta math. Soc., 35, 99. <https://ci.nii.ac.jp/naid/10030997340/en/>
- Biwer, C. M., Capano, C. D., De, S., et al. 2019, PASP, 131, 024503, doi: [10.1088/1538-3873/aaef0b](https://doi.org/10.1088/1538-3873/aaef0b)
- Blandford, R. D., & Begelman, M. C. 1999, Monthly Notices of the Royal Astronomical Society, 303, L1, doi: [10.1046/j.1365-8711.1999.02358.x](https://doi.org/10.1046/j.1365-8711.1999.02358.x)
- Blandford, R. D., & Znajek, R. L. 1977, MNRAS, 179, 433, doi: [10.1093/mnras/179.3.433](https://doi.org/10.1093/mnras/179.3.433)
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. 2017, Journal of the American Statistical Association, 112, 859, doi: [10.1080/01621459.2017.1285773](https://doi.org/10.1080/01621459.2017.1285773)
- Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- Bower, G. C., Goss, W. M., Falcke, H., Backer, D. C., & Lithwick, Y. 2006, ApJ, 648, L127, doi: [10.1086/508019](https://doi.org/10.1086/508019)
- Box, G. E. P. 1976, Journal of the American Statistical Association, 71, 791, doi: [10.1080/01621459.1976.10480949](https://doi.org/10.1080/01621459.1976.10480949)
- Broderick, A., & Blandford, R. 2004, MNRAS, 349, 994, doi: [10.1111/j.1365-2966.2004.07582.x](https://doi.org/10.1111/j.1365-2966.2004.07582.x)
- Broderick, A. E., Fish, V. L., Doeleman, S. S., & Loeb, A. 2011, The Astrophysical Journal, 735, 110. <http://stacks.iop.org/0004-637X/735/i=2/a=110>
- Broderick, A. E., Johannsen, T., Loeb, A., & Psaltis, D. 2014, ApJ, 784, 7, doi: [10.1088/0004-637X/784/1/7](https://doi.org/10.1088/0004-637X/784/1/7)

- Broderick, A. E., & Loeb, A. 2005, MNRAS, 363, 353, doi: [10.1111/j.1365-2966.2005.09458.x](https://doi.org/10.1111/j.1365-2966.2005.09458.x)
- . 2006, MNRAS, 367, 905, doi: [10.1111/j.1365-2966.2006.10152.x](https://doi.org/10.1111/j.1365-2966.2006.10152.x)
 - . 2009, ApJ, 703, L104, doi: [10.1088/0004-637X/703/2/L104](https://doi.org/10.1088/0004-637X/703/2/L104)
- Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020, ApJ, 898, 9, doi: [10.3847/1538-4357/ab9c1f](https://doi.org/10.3847/1538-4357/ab9c1f)
- Broderick, A. E., Tiede, P., Pesce, D. W., & Gold, R. 2021, Measuring Spin from Relative Photon Ring Sizes. <https://arxiv.org/abs/2105.09962>
- Broderick, A. E., Fish, V. L., Johnson, M. D., et al. 2016, ApJ, 820, 137. <http://stacks.iop.org/0004-637X/820/i=2/a=137>
- Broderick, A. E., Gold, R., Karami, M., et al. 2020, ApJ, 897, 139, doi: [10.3847/1538-4357/ab91a4](https://doi.org/10.3847/1538-4357/ab91a4)
- Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. 2011, Handbook of markov chain monte carlo (CRC press)
- Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, Journal of Statistical Software, Articles, 76, 1, doi: [10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01)
- Chael, A. 2019, PhD thesis, Harvard University
- Chael, A., Bouman, K., Johnson, M., et al. 2019, eht-imaging: v1.1.0: Imaging interferometric data with regularized maximum likelihood, doi: [10.5281/zenodo.2614016](https://doi.org/10.5281/zenodo.2614016)
- Chael, A. A. 2019, PhD thesis, Harvard University, Graduate School of Arts & Sciences
- Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018a, ApJ, 857, 23, doi: [10.3847/1538-4357/aab6a8](https://doi.org/10.3847/1538-4357/aab6a8)
- . 2018b, ApJ, 857, 23, doi: [10.3847/1538-4357/aab6a8](https://doi.org/10.3847/1538-4357/aab6a8)
- Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016a, ApJ, 829, 11, doi: [10.3847/0004-637X/829/1/11](https://doi.org/10.3847/0004-637X/829/1/11)
- . 2016b, ApJ, 829, 11, doi: [10.3847/0004-637X/829/1/11](https://doi.org/10.3847/0004-637X/829/1/11)

Choi, E., & Lee, C. 2003, Pattern Recognition, 36, 1703, doi: [10.1016/S0031-3203\(03\)00035-9](https://doi.org/10.1016/S0031-3203(03)00035-9)

Clark, B. G. 1980, A&A, 89, 377

Cornish, R., Vanetti, P., Bouchard-Côté, A., Deligiannidis, G., & Doucet, A. 2019, Scalable Metropolis-Hastings for Exact Bayesian Inference with Large Datasets. <https://arxiv.org/abs/1901.09881>

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

Das, S., & Suganthan, P. N. 2011, IEEE Transactions on Evolutionary Computation, 15, 4, doi: [10.1109/TEVC.2010.2059031](https://doi.org/10.1109/TEVC.2010.2059031)

Davelaar, J., Moscibrodzka, M., Bronzwaer, T., & Falcke, H. 2017, ArXiv e-prints. <https://arxiv.org/abs/1712.02266>

De, S., Finstad, D., Lattimer, J. M., et al. 2018, Phys. Rev. Lett., 121, 091102, doi: [10.1103/PhysRevLett.121.091102](https://doi.org/10.1103/PhysRevLett.121.091102)

Del Moral, P., Doucet, A., & Jasra, A. 2006, J Royal Statistical Soc B, 68, 411, doi: [10.1111/j.1467-9868.2006.00553.x](https://doi.org/10.1111/j.1467-9868.2006.00553.x)

Dexter, J., Kelly, B., Bower, G. C., et al. 2014, Monthly Notices of the Royal Astronomical Society, 442, 2797, doi: [10.1093/mnras/stu1039](https://doi.org/10.1093/mnras/stu1039)

Di Matteo, T., Johnstone, R. M., Allen, S. W., & Fabian, A. C. 2001, ApJ, 550, L19, doi: [10.1086/319489](https://doi.org/10.1086/319489)

Di Matteo, T., Quataert, E., Allen, S. W., Narayan, R., & Fabian, A. C. 2000, Monthly Notices of the Royal Astronomical Society, 311, 507, doi: [10.1046/j.1365-8711.2000.03134.x](https://doi.org/10.1046/j.1365-8711.2000.03134.x)

Di Valentino, E., Mena, O., Pan, S., et al. 2021, arXiv e-prints, arXiv:2103.01183. <https://arxiv.org/abs/2103.01183>

Dieleman, S., Willett, K. W., & Dambre, J. 2015, Mon Not R Astron Soc, 450, 1441, doi: [10.1093/mnras/stv632](https://doi.org/10.1093/mnras/stv632)

Dirac, P. A. M. 1950, Canadian journal of mathematics, 2, 129

Do, T., Ghez, A. M., Morris, M. R., et al. 2009, ApJ, 691, 1021, doi: [10.1088/0004-637X/691/2/1021](https://doi.org/10.1088/0004-637X/691/2/1021)

- Dovčiak, M., Karas, V., & Yaqoob, T. 2004, ApJS, 153, 205, doi: [10.1086/421115](https://doi.org/10.1086/421115)
- Duda, R. O., & Hart, P. E. 1972, Commun. ACM, 15, 11–15, doi: [10.1145/361237.361242](https://doi.org/10.1145/361237.361242)
- Eckart, A., Schödel, R., Meyer, L., et al. 2006, A&A, 455, 1, doi: [10.1051/0004-6361:20064948](https://doi.org/10.1051/0004-6361:20064948)
- Eckart, A., Baganoff, F. K., Morris, M., et al. 2004, A&A, 427, 1, doi: [10.1051/0004-6361:20040495](https://doi.org/10.1051/0004-6361:20040495)
- Eckart, A., Baganoff, F. K., Zamaninasab, M., et al. 2008a, A&A, 479, 625, doi: [10.1051/0004-6361:20078793](https://doi.org/10.1051/0004-6361:20078793)
- Eckart, A., Schödel, R., García-Marín, M., et al. 2008b, A&A, 492, 337, doi: [10.1051/0004-6361:200810924](https://doi.org/10.1051/0004-6361:200810924)
- Eckart, A., Baganoff, F. K., Morris, M. R., et al. 2009, A&A, 500, 935, doi: [10.1051/0004-6361/200811354](https://doi.org/10.1051/0004-6361/200811354)
- Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJ, 875, L1, doi: [10.3847/2041-8213/ab0ec7](https://doi.org/10.3847/2041-8213/ab0ec7)
- . 2019b, ApJ, 875, L2, doi: [10.3847/2041-8213/ab0c96](https://doi.org/10.3847/2041-8213/ab0c96)
- . 2019c, ApJ, 875, L3, doi: [10.3847/2041-8213/ab0c57](https://doi.org/10.3847/2041-8213/ab0c57)
- . 2019d, ApJ, 875, L4, doi: [10.3847/2041-8213/ab0e85](https://doi.org/10.3847/2041-8213/ab0e85)
- . 2019e, ApJ, 875, L5, doi: [10.3847/2041-8213/ab0f43](https://doi.org/10.3847/2041-8213/ab0f43)
- . 2019f, ApJ, 875, L6, doi: [10.3847/2041-8213/ab1141](https://doi.org/10.3847/2041-8213/ab1141)
- Event Horizon Telescope Collaboration, Akiyama, K., Algaba, J. C., et al. 2021a, ApJ, 910, L12, doi: [10.3847/2041-8213/abe71d](https://doi.org/10.3847/2041-8213/abe71d)
- . 2021b, ApJ, 910, L13, doi: [10.3847/2041-8213/abe4de](https://doi.org/10.3847/2041-8213/abe4de)
- Fazio, G. G., Hora, J. L., Witzel, G., et al. 2018, ApJ, 864, 58, doi: [10.3847/1538-4357/aad4a2](https://doi.org/10.3847/1538-4357/aad4a2)
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601, doi: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x)

Fish, V. L., Doebleman, S. S., Beaudoin, C., et al. 2011, ApJ, 727, L36, doi: [10.1088/2041-8205/727/2/L36](https://doi.org/10.1088/2041-8205/727/2/L36)

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: [10.1086/670067](https://doi.org/10.1086/670067)

Frieden, B. R. 1972, Journal of the Optical Society of America (1917-1983), 62, 511. <http://adsabs.harvard.edu/abs/1972JOSA...62..511F>

Gebhardt, K., Adams, J., Richstone, D., et al. 2011, ApJ, 729, 119, doi: [10.1088/0004-637X/729/2/119](https://doi.org/10.1088/0004-637X/729/2/119)

Gelman, A., & Rubin, D. B. 1992, Statistical Science, 7, 457. <http://www.jstor.org/stable/2246093>

Genzel, R., Schödel, R., Ott, T., & Eckart, A. 2003, Nature, 425, 7. <http://www.nature.com/nature/journal/v425/n6961/abs/nature02065.html>

Geyer, C. J. 1991, in Markov Chain Monte Carlo Maximum Likelihood (Interface Foundation of North America). <http://conservancy.umn.edu/handle/11299/58440>

Geyer, C. J. 1992, Statistical Science, 7, 473 , doi: [10.1214/ss/1177011137](https://doi.org/10.1214/ss/1177011137)

Ghez, A. M., Duchêne, G., Matthews, K., et al. 2003, ApJ, 586, L127, doi: [10.1086/374804](https://doi.org/10.1086/374804)

Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, ApJ, 689, 1044, doi: [10.1086/592738](https://doi.org/10.1086/592738)

Gillessen, S., Eisenhauer, F., Quataert, E., et al. 2006, Journal of Physics Conference Series, 54, 411, doi: [10.1088/1742-6596/54/1/065](https://doi.org/10.1088/1742-6596/54/1/065)

Goffe, W. L. 01 Oct. 1996, Studies in Nonlinear Dynamics & Econometrics, 1, doi: <https://doi.org/10.2202/1558-3708.1020>

Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press)

Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65, doi: [10.2140/camcos.2010.5.65](https://doi.org/10.2140/camcos.2010.5.65)

Goudail, F., Réfrégier, P., & Delyon, G. 2004, J. Opt. Soc. Am. A, 21, 1231, doi: [10.1364/JOSAA.21.001231](https://doi.org/10.1364/JOSAA.21.001231)

Gravity Collaboration, Abuter, R., Amorim, A., et al. 2018, A&A, 618, L10, doi: [10.1051/0004-6361/201834294](https://doi.org/10.1051/0004-6361/201834294)

- Grünwald, P., & van Ommen, T. 2018, Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. <https://arxiv.org/abs/1412.3730>
- Gull, S. F., & Daniell, G. J. 1978, Nature, 272, 686, doi: [10.1038/272686a0](https://doi.org/10.1038/272686a0)
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015, Monthly Notices of the Royal Astronomical Society: Letters, 450, L61–L65, doi: [10.1093/mnrasl/slv047](https://doi.org/10.1093/mnrasl/slv047)
- Hastings, W. K. 1970, Biometrika, 57, 97
- Hawking, S. W. 1976, Phys. Rev. D, 14, 2460, doi: [10.1103/PhysRevD.14.2460](https://doi.org/10.1103/PhysRevD.14.2460)
- Heckman, T. M., & Best, P. N. 2014, Annual Review of Astronomy and Astrophysics, 52, 589–660, doi: [10.1146/annurev-astro-081913-035722](https://doi.org/10.1146/annurev-astro-081913-035722)
- Heisenberg, W. 1938, Zeitschrift fur Physik, 110, 251, doi: [10.1007/BF01342872](https://doi.org/10.1007/BF01342872)
- Ho, L. C., Terashima, Y., & Ulvestad, J. S. 2003, ApJ, 589, 783, doi: [10.1086/374738](https://doi.org/10.1086/374738)
- Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints, arXiv:1111.4246. <https://arxiv.org/abs/1111.4246>
- Hogbom, J. A. 1974, Astronomy and Astrophysics Supplement Series, 15, 417. <http://adsabs.harvard.edu/abs/1974A%26AS...15..417H>
- Honma, M., Akiyama, K., Uemura, M., & Ikeda, S. 2014, PASJ, 66, 95, doi: [10.1093/pasj/psu070](https://doi.org/10.1093/pasj/psu070)
- Hooft, G. t. 1973, Nuclear Physics B, 62, 444, doi: [https://doi.org/10.1016/0550-3213\(73\)90263-0](https://doi.org/10.1016/0550-3213(73)90263-0)
- Hooft, G. t., & Veltman, M. J. G. 1974, Annales de l’Institut Henri Poincaré : Section A, Physique théorique, 20, 69. <http://localhost/handle/1874/4705>
- Hora, J. L., Witzel, G., Ashby, M. L. N., et al. 2014, ApJ, 793, 120, doi: [10.1088/0004-637X/793/2/120](https://doi.org/10.1088/0004-637X/793/2/120)
- Hough, P. V. C. 1964, Optical Engineering, 2, 379, doi: [10.1117/12.7971269](https://doi.org/10.1117/12.7971269)
- Huijser, D., Goodman, J., & Brewer, B. J. 2015, arXiv preprint arXiv:1509.02230
- Hukushima, K., & Nemoto, K. 1996, J. Phys. Soc. Jpn., 65, 1604, doi: [10.1143/JPSJ.65.1604](https://doi.org/10.1143/JPSJ.65.1604)

Ikeda, S., Tazaki, F., Akiyama, K., Hada, K., & Honma, M. 2016, PASJ, 68, 45, doi: [10.1093/pasj/psw042](https://doi.org/10.1093/pasj/psw042)

Innes, M. 2018, Journal of Open Source Software, doi: [10.21105/joss.00602](https://doi.org/10.21105/joss.00602)

Innes, M., Saba, E., Fischer, K., et al. 2018, CoRR, abs/1811.01457. <https://arxiv.org/abs/1811.01457>

Israel, W. 1967, Phys. Rev., 164, 1776, doi: [10.1103/PhysRev.164.1776](https://doi.org/10.1103/PhysRev.164.1776)

Issaoun, S., Johnson, M. D., Blackburn, L., et al. 2019, ApJ, 871, 30, doi: [10.3847/1538-4357/aaf732](https://doi.org/10.3847/1538-4357/aaf732)

—. 2021, arXiv e-prints, arXiv:2104.07610. <https://arxiv.org/abs/2104.07610>

Jeter, B., & Broderick, A. E. 2021, ApJ, 908, 139, doi: [10.3847/1538-4357/abda3d](https://doi.org/10.3847/1538-4357/abda3d)

Jeter, B., Broderick, A. E., & McNamara, B. R. 2019, The Astrophysical Journal, 882, 82, doi: [10.3847/1538-4357/ab3221](https://doi.org/10.3847/1538-4357/ab3221)

Johannsen, T. 2013, Phys. Rev. D, 88, 044002, doi: [10.1103/PhysRevD.88.044002](https://doi.org/10.1103/PhysRevD.88.044002)

Johannsen, T., & Psaltis, D. 2010, ApJ, 718, 446, doi: [10.1088/0004-637X/718/1/446](https://doi.org/10.1088/0004-637X/718/1/446)

Johannsen, T., & Psaltis, D. 2011, PRD, 83, doi: [10.1103/PhysRevD.83.124015](https://doi.org/10.1103/PhysRevD.83.124015)

Johannsen, T., Wang, C., Broderick, A. E., et al. 2016, Phys. Rev. Lett., 117, 091101, doi: [10.1103/PhysRevLett.117.091101](https://doi.org/10.1103/PhysRevLett.117.091101)

Johannsen, T., Broderick, A. E., Plewa, P. M., et al. 2016, Phys. Rev. Lett., 116, 031101, doi: [10.1103/PhysRevLett.116.031101](https://doi.org/10.1103/PhysRevLett.116.031101)

Johnson, M. D., Narayan, R., Psaltis, D., et al. 2018, ApJ, 865, 104, doi: [10.3847/1538-4357/aadcff](https://doi.org/10.3847/1538-4357/aadcff)

Johnson, M. D., Lupsasca, A., Strominger, A., et al. 2020, Science Advances, 6, eaaz1310, doi: [10.1126/sciadv.aaz1310](https://doi.org/10.1126/sciadv.aaz1310)

Kerr, R. P. 1963, Phys. Rev. Lett., 11, 237, doi: [10.1103/PhysRevLett.11.237](https://doi.org/10.1103/PhysRevLett.11.237)

Kim, J., Krichbaum, T. P., Broderick, A. E., et al. 2020, A&A, doi: [10.1051/0004-6361/202037493](https://doi.org/10.1051/0004-6361/202037493)

- Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980. <https://arxiv.org/abs/1412.6980>
- Kleijn, B., & van der Vaart, A. 2012, Electronic Journal of Statistics, 6, 354 , doi: [10.1214/EJS675](https://doi.org/10.1214/EJS675)
- Kullback, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79, doi: [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
- Kuramochi, K., Akiyama, K., Ikeda, S., et al. 2018, The Astrophysical Journal, 858, 56, doi: [10.3847/1538-4357/aab6b5](https://doi.org/10.3847/1538-4357/aab6b5)
- Luminet, J.-P. 1979, Astronomy and Astrophysics, 75, 228
- Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N., & Jordan, M. I. 2019, Proceedings of the National Academy of Sciences, 116, 20881, doi: [10.1073/pnas.1820003116](https://doi.org/10.1073/pnas.1820003116)
- Marrone, D. P. 2006, PhD thesis, Harvard University
- Marrone, D. P., Baganoff, F. K., Morris, M. R., et al. 2008, ApJ, 682, 373, doi: [10.1086/588806](https://doi.org/10.1086/588806)
- Mathur, S. D. 2009, Classical and Quantum Gravity, 26, 224001, doi: [10.1088/0264-9381/26/22/224001](https://doi.org/10.1088/0264-9381/26/22/224001)
- Medeiros, L., Psaltis, D., & Özel, F. 2020, ApJ, 896, 7
- Mertens, F., Lobanov, A. P., Walker, R. C., & Hardee, P. E. 2016, A&A, 595, A54, doi: [10.1051/0004-6361/201628829](https://doi.org/10.1051/0004-6361/201628829)
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087, doi: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114)
- Meyer, L., Do, T., Ghez, A., et al. 2009, ApJ, 694, L87, doi: [10.1088/0004-637X/694/1/L87](https://doi.org/10.1088/0004-637X/694/1/L87)
- Meyer, L., Witzel, G., Longstaff, F. A., & Ghez, A. M. 2014, ApJ, 791, 24, doi: [10.1088/0004-637X/791/1/24](https://doi.org/10.1088/0004-637X/791/1/24)
- Moriyama, K., Mineshige, S., Honma, M., & Akiyama, K. 2019, ApJ, 887, 227, doi: [10.3847/1538-4357/ab505b](https://doi.org/10.3847/1538-4357/ab505b)

- Mossoux, E., Grosso, N., Vincent, F. H., & Porquet, D. 2015, in SF2A-2015: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, 171–174
- Narayan, R., Mahadevan, R., Grindlay, J. E., Popham, R. G., & Gammie, C. 1998, ApJ, 492, 554, doi: [10.1086/305070](https://doi.org/10.1086/305070)
- Narayan, R., & Nityananda, R. 1986, Annual Review of Astronomy and Astrophysics, 24, 127, doi: [10.1146/annurev.aa.24.090186.001015](https://doi.org/10.1146/annurev.aa.24.090186.001015)
- Narayan, R., & Nityananda, R. 1986, ARA&A, 24, 127, doi: [10.1146/annurev.aa.24.090186.001015](https://doi.org/10.1146/annurev.aa.24.090186.001015)
- Neal, R. M. 2003, The Annals of Statistics, 31, 705 , doi: [10.1214/aos/1056562461](https://doi.org/10.1214/aos/1056562461)
- Neilsen, J., Nowak, M. A., Gammie, C., et al. 2013, ApJ, 774, 42, doi: [10.1088/0004-637X/774/1/42](https://doi.org/10.1088/0004-637X/774/1/42)
- Nelson, B., Ford, E. B., & Payne, M. J. 2014, ApJS, 210, 11, doi: [10.1088/0067-0049/210/1/11](https://doi.org/10.1088/0067-0049/210/1/11)
- Pesce, D. W. 2021, The Astronomical Journal, 161, 178, doi: [10.3847/1538-3881/abe3f8](https://doi.org/10.3847/1538-3881/abe3f8)
- Pompe, E., Holmes, C., Latuszyński, K., et al. 2020, Annals of Statistics, 48, 2930
- Ponti, G., George, E., Scaringi, S., et al. 2017, MNRAS, 468, 2447, doi: [10.1093/mnras/stx596](https://doi.org/10.1093/mnras/stx596)
- Porquet, D., Grosso, N., Predehl, P., et al. 2008, A&A, 488, 549, doi: [10.1051/0004-6361:200809986](https://doi.org/10.1051/0004-6361:200809986)
- Prieto, M. A., Fernández-Ontiveros, J. A., Markoff, S., Espada, D., & González-Martín, O. 2016, Monthly Notices of the Royal Astronomical Society, 457, 3801–3816, doi: [10.1093/mnras/stw166](https://doi.org/10.1093/mnras/stw166)
- Psaltis, D., Özel, F., Chan, C.-K., & Marrone, D. P. 2015, ApJ, 814, 115, doi: [10.1088/0004-637X/814/2/115](https://doi.org/10.1088/0004-637X/814/2/115)
- Psaltis, D., Wex, N., & Kramer, M. 2016, ApJ, 818, 121, doi: [10.3847/0004-637X/818/2/121](https://doi.org/10.3847/0004-637X/818/2/121)
- Pu, H.-Y., Akiyama, K., & Asada, K. 2016, ApJ, 831, 4, doi: [10.3847/0004-637X/831/1/4](https://doi.org/10.3847/0004-637X/831/1/4)
- Rees, M. J. 1984, ARA&A, 22, 471, doi: [10.1146/annurev.aa.22.090184.002351](https://doi.org/10.1146/annurev.aa.22.090184.002351)

- Reynolds, C. 2003, Physics Reports, 377, 389–466, doi: [10.1016/s0370-1573\(02\)00584-7](https://doi.org/10.1016/s0370-1573(02)00584-7)
- Reynolds, C. S., Di Matteo, T., Fabian, A. C., Hwang, U., & Canizares, C. R. 1996, MNRAS, 283, L111, doi: [10.1093/mnras/283.4.L111](https://doi.org/10.1093/mnras/283.4.L111)
- Ripperda, B., Bacchini, F., & Philippov, A. A. 2020, The Astrophysical Journal, 900, 100, doi: [10.3847/1538-4357/ababab](https://doi.org/10.3847/1538-4357/ababab)
- Robert, C., & Casella, G. 2013, Monte Carlo statistical methods (Springer Science & Business Media)
- Roberts, G. O., & Rosenthal, J. S. 2001, Statistical Science, 16, 351 , doi: [10.1214/ss/1015346320](https://doi.org/10.1214/ss/1015346320)
- . 2009, Journal of Computational and Graphical Statistics, 18, 349, doi: [10.1198/jcgs.2009.06134](https://doi.org/10.1198/jcgs.2009.06134)
- Rosenfeld, L. 1930, Annalen der Physik, 397, 113
- Schmidt, M. 1963, Nature, 197, 1040, doi: [10.1038/1971040a0](https://doi.org/10.1038/1971040a0)
- Schödel, R., Ott, T., Genzel, R., et al. 2002, Nature, 419, 694, doi: [10.1038/nature01121](https://doi.org/10.1038/nature01121)
- Schwab, F. R. 1984, AJ, 89, 1076, doi: [10.1086/113605](https://doi.org/10.1086/113605)
- Schwarz, U. J. 1978, Astronomy and Astrophysics, 65, 345. <http://adsabs.harvard.edu/abs/1978A%26A....65..345S>
- Seyfert, C. K. 1943, ApJ, 97, 28, doi: [10.1086/144488](https://doi.org/10.1086/144488)
- Skilling, J. 2006, Bayesian Analysis, 1, 833 , doi: [10.1214/06-BA127](https://doi.org/10.1214/06-BA127)
- Sokal, A. 1996, in Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms Note to the Reader
- Stefanov, I. Z., Yazadjiev, S. S., & Gyulchev, G. G. 2010, Phys. Rev. Lett., 104, 251103, doi: [10.1103/PhysRevLett.104.251103](https://doi.org/10.1103/PhysRevLett.104.251103)
- Sun, H., & Bouman, K. L. 2020, arXiv preprint arXiv:2010.14462
- Swendsen, R. H., & Wang, J.-S. 1986, Phys. Rev. Lett., 57, 2607, doi: [10.1103/PhysRevLett.57.2607](https://doi.org/10.1103/PhysRevLett.57.2607)

- Syed, S., Bouchard-Côté, A., Deligiannidis, G., & Doucet, A. 2019, arXiv:1905.02939 [stat].
<http://arxiv.org/abs/1905.02939>
- Syed, S., Romaniello, V., Campbell, T., & Bouchard-Côté, A. 2021, arXiv:2102.07720 [stat]. <http://arxiv.org/abs/2102.07720>
- Teo, E. 2003, General Relativity and Gravitation, 35, 1909
- Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition (Springer International Publishing), doi: [10.1007/978-3-319-44431-4](https://doi.org/10.1007/978-3-319-44431-4)
- Tibbits, M. M., Groendyke, C., Haran, M., & Liechty, J. C. 2014, Journal of Computational and Graphical Statistics, 23, 543, doi: [10.1080/10618600.2013.791193](https://doi.org/10.1080/10618600.2013.791193)
- Tiede, P., & Broderick, A. E. in prep., Bayesian Ellipticity Estimates of M 87
- Tiede, P., Broderick, A. E., & Palumbo, D. C. M. 2020, arXiv:2012.07889, Variational Image Feature Extraction for the EHT
- Tiede, P., Broderick, A. E., Palumbo, D. C. M., & Chael, A. in prep.a, Measuring the Ellipticity of the M 87 EHT Images
- Tiede, P., Broderick, A. E., & Saifuddin, S. in prep.b, Next Generation Samplers for THEMIS
- Tiede, P., Pu, H.-Y., Broderick, A. E., et al. 2020, The Astrophysical Journal, 892, 132.
<https://doi.org/10.3847/1538-4357/ab744c>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2021, Bayesian Analysis, -1, doi: [10.1214/20-ba1221](https://doi.org/10.1214/20-ba1221)
- Vousden, W. D., Farr, W. M., & Mandel, I. 2016, MNRAS, 455, 1919, doi: [10.1093/mnras/stv2422](https://doi.org/10.1093/mnras/stv2422)
- Walsh, J. L., Barth, A. J., Ho, L. C., & Sarzi, M. 2013, ApJ, 770, 86, doi: [10.1088/0004-637X/770/2/86](https://doi.org/10.1088/0004-637X/770/2/86)
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009a, MNRAS, 395, 1733, doi: [10.1111/j.1365-2966.2009.14665.x](https://doi.org/10.1111/j.1365-2966.2009.14665.x)
- Wiaux, Y., Puy, G., Boursier, Y., & Vandergheynst, P. 2009b, MNRAS, 400, 1029, doi: [10.1111/j.1365-2966.2009.15519.x](https://doi.org/10.1111/j.1365-2966.2009.15519.x)

- Wielgus, M., Akiyama, K., Blackburn, L., et al. 2020, ApJ, 901, 67, doi: [10.3847/1538-4357/abac0d](https://doi.org/10.3847/1538-4357/abac0d)
- Witten, E. 1988, Communications in Mathematical Physics, 117, 353, doi: [10.1007/BF01223371](https://doi.org/10.1007/BF01223371)
- Witzel, G., Eckart, a., Bremer, M., et al. 2012, ApJ Suppl. Series, 203, 18, doi: [10.1088/0067-0049/203/2/18](https://doi.org/10.1088/0067-0049/203/2/18)
- Witzel, G., Martinez, G., Hora, J., et al. 2018, ApJ, 863, 15, doi: [10.3847/1538-4357/aace62](https://doi.org/10.3847/1538-4357/aace62)
- Wong, G. N. 2021, The Astrophysical Journal, 909, 217, doi: [10.3847/1538-4357/abdd2d](https://doi.org/10.3847/1538-4357/abdd2d)
- Yang, H. 2021, Phys. Rev. D, 103, 084010, doi: [10.1103/PhysRevD.103.084010](https://doi.org/10.1103/PhysRevD.103.084010)
- Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. 2018, in Proceedings of Machine Learning Research, Vol. 80, Proceedings of the 35th International Conference on Machine Learning, ed. J. Dy & A. Krause (PMLR), 5581–5590. <http://proceedings.mlr.press/v80/yao18a.html>
- Yuan, F., Markoff, S., & Falcke, H. 2002, A&A, 383, 854, doi: [10.1051/0004-6361:20011709](https://doi.org/10.1051/0004-6361:20011709)
- Yuan, F., & Narayan, R. 2014, ARA&A, 52, 529, doi: [10.1146/annurev-astro-082812-141003](https://doi.org/10.1146/annurev-astro-082812-141003)
- Yuan, F., Quataert, E., & Narayan, R. 2004, ApJ, 606, 894, doi: [10.1086/383117](https://doi.org/10.1086/383117)
- Zamaninasab, M., Eckart, A., Witzel, G., et al. 2010, A&A, 510, A3, doi: [10.1051/0004-6361/200912473](https://doi.org/10.1051/0004-6361/200912473)

Appendices

Appendix A

Complex Gain Marginalization

In the ideal case the complex visibilities V_{ij} are related to the on-sky image through the two dimensional Fourier transform. Unfortunately, due to atmospheric and telescope effects these complex visibilities are modified by complex numbers g_i for each station, called *complex gains*. The measured quantities are then given by

$$\tilde{V}_{ij} = g_i g_j^* V_{ij} = \gamma_i \gamma_j \exp[i(\theta_i - \theta_j)], \quad (\text{A.1})$$

where $\gamma_{i,j}$ and $\phi_{i,j}$ are known as the gain amplitude and phases respectively. While prepossessing of the EHT data can reduce the impact of the gains, they still can affect the measured visibility amplitudes by 10%, and can extend up to 100% for certain stations such as LMT. Furthermore, the gain phases are basically scrambled by atmosphere. Therefore, these gains form an unknown or latent parameter that should be forward modeled in any Bayesian analysis. Unfortunately, the gains are time dependent and can change drastically from scan to scan due to, e.g., pointing issues. Therefore, directly modeling the gains can introduce hundreds of additional parameters into the forward model, drastically increasing the computational timescale.

However, Bayesian inference provides a way to deal with these parameters that can effectively reduce the total number of parameters we need to model through marginalization. To see this first note that the joint distribution for a static model θ with gains \mathbf{g} can be written as

$$p(\mathbf{V}, \mathbf{g}, \theta) = p(\theta) \prod_s p(\mathbf{V}_s | \mathbf{g}_s, \theta) p(\mathbf{g}_s), \quad (\text{A.2})$$

where s denotes the scan over which the gains are assumed to be constant¹, and \mathbf{V}_s and \mathbf{g}_s

¹For the 2017 EHT data this was typically set to be approximately 12 minutes.

are the complex visibilities measured and station gains during said scan.

$$p(\theta, \mathbf{g} | \mathbf{V}) = \frac{1}{p(\mathbf{V})} p(\theta) \prod_s p(\mathbf{V}_s | \theta, \mathbf{g}_s) p(\mathbf{g}_s). \quad (\text{A.3})$$

The gains then essentially function as nuisance parameters needed to fully model the observations, but are of little physical interest. Therefore, we will attempt to marginalize over the gains, i.e. our posterior will be

$$p(\theta | \mathbf{V}) = \frac{1}{p(\mathbf{V})} p(\theta) \prod_s \int p(\mathbf{V}_s | \theta, \mathbf{g}_s) p(\mathbf{g}_s) d\mathbf{g}_s. \quad (\text{A.4})$$

Unfortunately, this integral isn't analytically tractable. While there exist a number of Monte Carlo methods to evaluate integrals of this form (e.g. Sequential Monte Carlo (Del Moral et al., 2006)), we will instead use the Laplace approximation. The Laplace approximation is a simple version of the saddle-point approximation used in physics to evaluate integrals of the form $\int \exp(f(x))dx$. The first step is to maximize the function in the integrand of Equation A.4. Namely we seek the complex gains $\hat{\mathbf{g}}_s$ such that $p(\mathbf{V}_s | \theta, \mathbf{g}_s) p(\mathbf{g}_s)$ is maximized. Given this point we will then Taylor expand the logarithm of the integrand around $\hat{\mathbf{g}}_s$, and truncate at second order:

$$\int p(\mathbf{V}_s | \theta, \mathbf{g}_s) p(\mathbf{g}_s) \approx p(\mathbf{V}_s | \theta, \hat{\mathbf{g}}_s) p(\hat{\mathbf{g}}_s) \int \exp \left[\frac{1}{2} (\mathbf{g}_s - \hat{\mathbf{g}}_s)^\top \mathbf{C}_s (\mathbf{g}_s - \hat{\mathbf{g}}_s) \right], \quad (\text{A.5})$$

where $\mathbf{C}_s = \frac{\partial \log \mathcal{L}}{\partial \mathbf{g}_s \partial \mathbf{g}_s^\top} \Big|_{\hat{\mathbf{g}}}$. Note that due to $\hat{\mathbf{g}}$ being a maximum \mathbf{C}_s is a negative definite matrix. (A.4) is now a simple Gaussian integral giving:

$$p(\theta | \mathbf{V}) = \frac{p(\theta)}{p(\mathbf{V})} \prod_s \sqrt{-\det \mathbf{C}_s} (2\pi)^{N_s} \mathcal{L}(\mathbf{V}_s | \theta, \hat{\mathbf{g}}_s) p(\hat{\mathbf{g}}_s), \quad (\text{A.6})$$

where N_s is the number of stations during scan s . This defines the marginal posterior distribution used in all THEMIS analyses when fitting complex gains. Typically we choose $p(\mathbf{g}) = p(\boldsymbol{\gamma})p(\boldsymbol{\phi})$, i.e. we fit the gain amplitudes and phases. For the gain amplitudes the prior is typically chosen to be log-normal with mean 0, i.e. $\gamma = 1$, and standard deviation of 0.1. The phase prior is either uniform or normal with a large standard deviation to reflect the potential scrambling effects of the atmosphere. With this choice of priors the gain optimization of the integrand in Equation A.4 becomes a least-squares optimization problem with a block diagonal structure. For the details of the optimization algorithm see Broderick et al. (2020).

Appendix B

Bayesian Information Criterion

In Bayesian analysis the model itself needs to be evaluated and considered during inference. One way to assess models is to use the evidence or marginal likelihood to “score” the relative performance of two competing models. However, computing the evidence for each model is quite expensive in practise. Instead approximations are often employed. One common approximation is the Bayesian Information Criterion (BIC) given by

$$\text{BIC} = -2 \log p(V|\hat{\theta})p(\hat{\theta}) + N_\theta \log N_v, \quad (\text{B.1})$$

where $\hat{\theta}$ is the best fit model, i.e. the maximum likelihood estimate, N_θ is the number of parameters, and N_V is the number of data points.

To derive the BIC we start with the Bayesian evidence:

$$p(V) = \int p(V|\theta)p(\theta)d\theta \quad (\text{B.2})$$

This integral is typically intractable. While the tempering methods mentioned in [Chapter 4](#) can compute it with thermodynamic integration, we can also approximate the integral using a Laplace approximation as in [Appendix B](#). This means that we approximation the logarithm of the integrand as:

$$\log p(V|\theta)p(\theta) \approx \log p(V|\hat{\theta})p(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^T H_{\hat{\theta}}(\theta - \hat{\theta}), \quad (\text{B.3})$$

where $\hat{\theta}$ is the maximum of $\log p(V|\theta)p(\theta)$ and $H_{\hat{\theta}}$ is the Hessian matrix around the peak, given by

$$H_{\hat{\theta}} = -\partial_\theta \partial_\theta \log p(V|\theta)p(\theta). \quad (\text{B.4})$$

Plugging this expansion into the evidence integral we get the log evidence is

$$\log p(V) = \log p(V|\hat{\theta})p(\hat{\theta}) + \frac{N_\theta}{2} \log 2\pi - \frac{1}{2} \log \det H_{\hat{\theta}}. \quad (\text{B.5})$$

To derive the BIC we need to make a number of assumptions:

1. The priors are rather flat near the maximum so we can ignore their contribution in the Hessian.
2. The number of data points is "large" and the $\mathbf{V} = (V_1, \dots, V_{N_v})$ are independently and identically distributed.

These assumptions allow us to greatly simplify the Hessian [Equation B.4](#)

$$\begin{aligned} H_{\hat{\theta}} &\approx -\partial_\theta \partial_{\theta'} \log p(\mathbf{V}|\theta) \quad (\text{using assumption 1.}) \\ &= -\partial_\theta \partial_{\theta'} \sum_{i=1}^N \log p(V_i|\theta) \\ &= \frac{1}{N_v} \sum_j -N_V \partial_\theta \partial_{\theta'} \log p(y_j|\theta) \\ &\approx E[N_V \log p(V_j|\theta)] \quad (\text{using assumption 2.}) \\ &= N_v I_{\theta\theta'}, \end{aligned}$$

where $I_{\theta\theta'}$, is known as the Fischer information matrix. The determinant of the Hessian is then given by:

$$\det H_{\hat{\theta}} \approx N_v^{N_\theta} \det I_{\theta\theta'}. \quad (\text{B.6})$$

With this approximation we the log-evidence is given by,

$$\log p(V) \approx \log p(V|\hat{\theta})p(\hat{\theta}) + \frac{N_\theta}{2} \log 2\pi - \frac{1}{2} N_\theta \log N_v - \frac{1}{2} \log \det I_{\theta\theta'}. \quad (\text{B.7})$$

In the large N_V limit we then get:

$$\log p(V) \approx \log p(V|\hat{\theta})p(\hat{\theta}) - \frac{1}{2} N_\theta \log N_v. \quad (\text{B.8})$$

The BIC is then defined as $-2p(V)$ and we get

$$\text{BIC} = -2 \log p(V|\hat{\theta})p(\hat{\theta}) + N_\theta \log N_v. \quad (\text{B.9})$$

Note that in the presence of gains $N_\theta \log N_v$ expression becomes more complicated since they are not "global parameters". However, we only care about the difference in the BIC between two models, and since the number of gains is constant, this contribution will drop out.