Extreme Imaging via Physical Model Inversion: Seeing Around Corners and Imaging Black Holes

by

Katherine L. Bouman

B.S.E., Electrical Engineering, University of Michigan, 2011
S.M., Electrical Engineering and Computer Science, M.I.T., 2013

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September 2017

Signature of Author: Signature redacted
Department of Electrical Engineering and Computer Science
August 31, 2017

Certified by: Signature redacted
Professor William T. Freeman
Thomas and Gerd Perkins Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: Signature redacted
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Committee for Graduate Students

Extreme Imaging via Physical Model Inversion: Seeing Around Corners and Imaging Black Holes

by Katherine L. Bouman

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

Imaging often plays a critical role in advancing fundamental science. However, as science continues to push the boundaries of knowledge, imaging systems are reaching the limits of what can be measured using traditional-direct approaches. By designing systems that tightly integrate novel sensor and algorithm design, it may be possible to develop imaging systems that exceed fundamental theoretical limitations to observe things previously impossible to see. However, these non-traditional imaging systems generally come with a trade-off; they produce increasingly sparse and/or noisy measurements that require incorporating additional structure to extract anything meaningful. The focus of this thesis is on using computational methods that exploit structure in our universe to move past these obstacles and reveal the invisible.

In this thesis, we focus on two imaging problems that explicitly leverage structure in our universe: reconstructing images and video from a computational telescope the size of the Earth, and seeing around corners. For the first imaging problem, this thesis investigates ways to reconstruct images and video from a sparse telescope array distributed around the globe. Additionally, it presents a number of evaluation techniques developed to rigorously evaluate imaging methods in order to establish confidence in reconstructions done with real scientific data. The methods and evaluation techniques developed in this thesis will hopefully aid in ongoing work to take the first picture of a black hole. Next, this thesis presents methods developed for using the subtle spatio-temporal radiance variations that arise on the ground at the base of an edge to construct a one-dimensional video of a hidden scene. These methods may be especially valuable in remotely sensing occupants in a room during search and rescue operations, or in detecting hidden, oncoming vehicles and/or pedestrians for collision avoidance systems.

Thesis Supervisor: Professor William T. Freeman

Title: Thomas and Gerd Perkins Professor of Electrical Engineering and Computer Science

Acknowledgments

With these next few paragraphs, I want to express my gratitude to the many people who have made my time at MIT an exceptional experience.

I would first like to thank my advisor, Bill Freeman, for his never-ending support and guidance. Bill has been a wonderful mentor. He always encouraged me to follow ideas I was motivated by and gave me the freedom to pursue research that I found exhilarating, even if it was far from normal for a computer vision Ph.D student. His constant curiosity and desire to gain insight from his work has taught me to strive for a deeper understanding of the problems and methods I work on. I always like to say that Bill is really an artist at heart. Although I somewhat mean it as a joke, I believe there is a lot of truth to this statement. Bill sees the world around us in a totally different way. When he goes outside he might look at some seemingly mundane, like how light hits a leaf, and figure out a way to do something totally ridiculous that I would have never thought of. His incredible creativity combined with mathematical rigor has been such an inspiration to me, and I strive to one day have just an ounce of the creativity he shows in solving problems.

I would also like the other members of my committee: Shep Doeleman and Polina Golland. In my first meeting with Shep on the Event Horizon Telescope project I understood essentially nothing of what he was saying. However, I left that meeting knowing that the EHT was something I wanted to work on. The reason why wasn't because the project was on black holes, that was kind of cool, but what really drew me was Shep's excitement. His passion made it clear to me that, no matter what, I would have fun working on the project. This has definitely been true every step of the way. Shep has been incredibly welcoming of me – treating me as a member of their tight knit group even though I started the project as an outsider. I thank him for all his support and the opportunities he has provided for me. I was fortunate to work with Polina on a project reconstructing high-resolution MRI brain scans. Whenever we would be really

confused about how to solve a problem we would talk to Polina and somehow she would make the solution much more clear. Polina doesn't let you get away with hand waving explanations – I would always have to justify the choices I make with her. I know that this has made me a much better researcher.

I am thankful to the many people that have created the lively, and productive environment I've had the pleasure to work in, both in the Vision Graphics Neighborhood at MIT and the Event Horizon Telescope. I would especially like to thank (in no particular order) Ramesh Sridharan, Tianfan Xue, Adrian Dalca, Vickie Ye, Miki Rubinstein, Phillip Isola, Andrew Owens, Carl Vondrick, Michael Johnson, Andrew Chael, Lindy Blackburn, Katherine Rosenfeld, George Chen, Zoya Bylinskii, Jiajun Wu, Xiuming Zhang, Randi Cabezas, Danielle Pace, Joseph Lim, Donglai Wei, Roger Grosse, Hossein Mobahi, Tali Dekel, Daniel Zoran, Dilip Krishnan, Oren Freifeld, Aditya Khosla, Bolei Zhou, Guy Rosman, David Hayden, Julian Straub, Neal Wadhwa, Sudeep Pillai, Nikhil Naik, Christian Wachinger, Andreea Bobu, Ruizhi Liao, Polina Binder, Miaomiao Zhang, Giorgos Papachristoudis, Mazy Abulnaga, Yichang Shih, Kazu Akiyama, Vincent Fish, Jim Moran, Hotaka Shiokawa, Emma Alexander, Hamed Pirsiavash, Michael Gharbi, Jonathan Weintroub, Andre Young, Atish Kamble, Gopal Narayanan, Freek Roelofs, Jason Chang, Michael Janssen, Geoff Crew, Lia Medeiros, Dimitrios Psaltis, Aleksandar Popstefanija, Rurik Primiani, Laura Vertatschitsch, Fredo Durand, Adam Yedidia, Ce Liu, CK Chan, Justin Chen, Antonio Torralba, Peter Battaglia, Bei Xiao, Edward Adelson, Ruth Rosenholtz, Wojciech Matusik, John Fisher, and Berthold Horn.

I want to particularly thank four people from this list: Michael Johnson, Andrew Chael, Lindy Blackburn, and Adrian Dalca. The first time I met Michael, Andrew and Lindy was at the EHT Meeting in Waterloo, and we all went together to see *Interstellar* (amazingly...). All three of them have been so welcoming of me to the EHT group and have been incredibly fun to work with, not only as collaborators, but also as friends. Michael and Andrew have been my fellow imaging buddies and Lindy taught me so much about telescopes and VLBI during our time in Mexico at the Large Millimeter Telescope. I also want to thank Adrian Dalca, who was the first to welcome me into VGN when I joined nearly six years ago. I couldn't have asked for a better friend and mentor than Adrian. Often times discussing with Adrian for just a few minutes will bring my mistakes to light (personal or academic), and a solution becomes so much more clear. I hope that even as we go on our separate ways we will find a way to continue working together.

I want to thank my family for all their love and support. Especially my parents, who fought for me to succeed even when many people didn't necessarily think I could. I didn't appreciate this when I was younger, but I know that I wouldn't be where I am now if it wasn't for what they did for me back then.

And lastly I want to thank my fiancé, Joseph Leong. Joe has been with me every step of the way – through the ups and downs of grad school he has been by my side. He has been my pillar of support, and I know I owe a lot of my success to him.

Thank You.

Contents

Abstract **3**

Acknowledgments **5**

List of Figures **15**

1 Introduction **35**

1.0.1 Imaging with an Earth-Sized Computational Telescope 36

1.0.2 Imaging Hidden Scenes Behind Corners 37

1.1 Inverse Problems 38

Forward Model 39

Prior 41

1.1.1 Bayesian Inference 41

Gaussian Distributions 42

1.1.2 Bias-Variance Trade-off 44

2 Radio Interferometry and the Event Horizon Telescope **45**

2.1 Single Dish Telescope 47

2.1.1 Power Response 47

2.1.2 Angular Resolution 49

2.2 Radio Interferometry 49

The van Cittert-Zernike Theorem and Visibility Measurements . 50

2.2.1 Intuition Under a Simple 1D Interferometer 50

Response of a Point Source 51

Response of an Extended Emission 54

The 1D van Cittert-Zernike Theorem 55

2.2.2 Polarization 56
2.2.3 Earth Rotation Synthesis 57
2.3 Very Long Baseline Interferometry (VLBI) 57
2.3.1 Instrument Noise & Calibration Errors 58
Thermal Noise 59
Atmospheric Propagation Error 59
Systematic Miscalibration Gain Error 60
2.3.2 Data Products 60
Complex Visibility 61
Visibility Amplitude 61
Bispectrum & Closure Phase 61
Closure Amplitude 62
2.4 The Event Horizon Telescope (EHT) 63
2.4.1 Testing General Relativity 63
2.4.2 Targeted Black Holes 64
Sagittarius A* (Sgr A*) 64
Messier 87 (M87) 65
2.4.3 Observing Conditions & Equipment 65
Observing Wavelength 65
Telescopes and uv-Coverage 66
Receiver and Backend Pipeline 67
2.4.4 Black Holes 67
The Schwarzschild Radius and No Hair Theorem 68
Accretion Disk, Inclination, and Doppler Beaming 68
Gravitational Lensing of the Last Photon Orbit 69
Time-Variability 71
2.4.5 Preliminary Results 72

3 Single Image VLBI Reconstruction of Static Sources 75
3.1 Previous Interferometric Imaging Approaches 76
3.1.1 CLEAN 77
Procedure 77
Self-Calibration 77
Discussion 78
3.1.2 Bayesian-Style Methods 79

Optical Interferometry . . . 80
3.2 Defining the Forward Model & Conditional Distribution . . . 80
3.2.1 Deterministic Forward Model . . . 81
Continuous Image Representation . . . 81
Extracting a Visibility . . . 83
Extracting Additional Data Products . . . 84
3.2.2 Conditional Distribution . . . 85
Gaussian Noise Approximation for the Bispectrum . . . 86
3.3 Static Imaging using Data-Driven Priors . . . 86
3.3.1 Gauss: Data-Driven Multivariate Gaussian Image Priors . . . 87
Model . . . 87
Training a Multivariate Gaussian Image Prior . . . 88
Optimization . . . 88
3.3.2 CHIRP: Data-Driven GMM Patch Priors . . . 90
Model . . . 90
Training a GMM Patch Prior . . . 91
Optimization . . . 91
3.4 Results and Discussion . . . 94
3.4.1 Synthetic Measurements . . . 94
3.4.2 Bias Introduced Under Trained Image & Patch Priors . . . 95
Training . . . 96
Visualizing the Models . . . 96
Comparing Model Bias . . . 98
3.4.3 Method Comparison . . . 103
Synthetic Data . . . 103
Real Measurements . . . 105
3.5 Conclusion . . . 107

4 Video VLBI Reconstruction of Time-Varying Sources 109
4.1 Alternate Approaches . . . 110
4.2 Dynamic Imaging Model . . . 112
4.2.1 Multivariate Gaussian Image Prior . . . 112
4.2.2 Evolution Model . . . 114
4.3 Inference & Learning . . . 117
4.3.1 Known Evolution . . . 117

4.3.2 Unknown Evolution 119
4.4 Results 122
4.4.1 Data Generation 122
4.4.2 Static Evolution Model (No Warp) 124
4.4.3 Unknown Evolution Model (Learn Warp) 126
4.5 Conclusion 130

5 Reconstruction Quality Evaluation & the EHT Imaging Challenges 133
5.1 VLBI Reconstruction Dataset 134
5.1.1 Synthetic Measurements 134
Generate Your Own Data 134
5.1.2 Real Measurements 135
5.1.3 Automatic Evaluation System 135
5.2 Event Horizon Telescope Imaging Challenges 135
5.2.1 Challenge Pipeline 137
5.2.2 Synthetic Data Generation 138
Selecting the True Image or Video 138
Generating EHT Visibilities 139
Adding Noise to the Visibilities 139
5.2.3 Analysis and Evaluation 140
Quantitative Analysis 140
Qualitative Analysis 142
5.2.4 Challenge Datasets & Results 143
Challenge 1 - Dataset 1 144
Challenge 1 - Dataset 2 147
Challenge 1 - Dataset 3 150
Challenge 2 - Dataset 4 153
Challenge 2 - Dataset 5 156
Challenge 2 - Dataset 6 159
Challenge 3 - Dataset 7 163
Challenge 4 - Dataset 8 167
Challenge 4 - Dataset 9 171
Challenge 4 - Dataset 10 175
5.2.5 Discussion 179
Encouraging Algorithm Development 179

Identifying Biases . . . 181
Identifying Degeneracies . . . 182
Identifying Masquerading Structures . . . 183
Beginning a Conversation About Confidence . . . 184

6 Turning Corners into Cameras **185**
6.1 Related Work . . . 187
6.2 Edge Cameras . . . 188
6.2.1 Method . . . 190
Implementation Details . . . 191
6.2.2 Experiments and Results . . . 193
Environments . . . 194
Video Quality: . . . 194
Velocity Estimation . . . 195
6.3 Stereo Edge Cameras . . . 195
6.3.1 Method . . . 195
6.3.2 Experiments and Results . . . 198
6.3.3 Error Analysis . . . 198
6.4 Conclusion . . . 201

A Static VLBI Imaging Derivations **203**
A.1 Bayesian Inference under Gaussian Distributions . . . 203
A.2 Interstellar Scattering . . . 204
A.3 Approximate Gaussian Noise on the Bispectrum . . . 205
A.4 Data Product Derivatives . . . 206
A.4.1 Bispectrum . . . 206
A.4.2 Visibility Amplitudes . . . 207

B StarWarps Derivations **209**
B.1 Solving for the $E[x_t]$ and $E[x_t x_t^T]$ Sufficient Statistics . . . 209
B.1.1 Forward Messages . . . 211
B.1.2 Backward Messages . . . 212
B.1.3 Putting it Together . . . 212
B.2 Solving for the $E[x_{t-1} x_t^T]$ Sufficient Statistic . . . 213
B.3 Likelihood . . . 215
B.4 Non-linear Observation Model . . . 217

B.4.1 Gamma Regularization 217
B.5 Estimating the Evolution Parameters 218

C Corner Camera Derivations **221**
C.1 Corner Location Errors 221
C.1.1 Edge Camera 221
C.1.2 Stereo Camera 225
C.2 Temporal Smoothing 228
C.2.1 Kalman Filtering/Smoothing 228

D General Lemmas **231**
D.0.1 Lemma 1 231
D.0.2 Lemma 2 231
D.0.3 Lemma 3 232
D.0.4 Lemma 4 232
D.0.5 Lemma 5 232
D.0.6 Lemma 6 233

Bibliography **235**

List of Figures

1.1 A traditional approach to solving an inverse problem might look something like what is illustrated in this diagram. The sensing system produces measurements $\boldsymbol{y}$ in responsc to thc unknown signal $\boldsymbol{x}$. To recover the unknown signal an inverse function is constructed, $f^{-1}(\boldsymbol{y})$, which acts on the observed measurements, $\boldsymbol{y}$, to produce an estimate of the unknown signal, $\hat{\boldsymbol{x}}$. Although this approach seems fairly straightforward, it has a number of flaws and often results in poor reconstructions. 39

1.2 In this thesis a Bayesian style inference approach is taken to solve inverse problems. The sensing system produces measurements $\boldsymbol{y}$ in response to the unknown signal $\boldsymbol{x}$. To recover the unknown signal we construct a function$f(\boldsymbol{x})$ which models the behavior of the system. We then search for a signal $\hat{\boldsymbol{x}}$ that would produce measurements, $\hat{\boldsymbol{y}}$, similar to what we have observed and also fits our model of likely behavior of the unknown signal, $\boldsymbol{x}$. 40

2.1 Images from the nearest radio galaxy, Centaurus A. (a) Optical Image, (b) Radio Image, (c) Composite Image. Notice that the galactic-sized radio lobes that are being emitted from the supermassive black hole at the center of the galaxy are invisible to the optical image (a), and can only be seen when imaged in the radio (b). [76] 45

2.2 A single-dish telescope uses a parabolic mirror to reflect and coherently collect light at a focal point. Due to the geometry of a parabola, light coming from a coherent wave front traveling parallel to the axis of the dish, will travel an equal distance to the focus from every reflected point (a). Thus, the electric field will be in phase, and constructively interference, at the focus. When the wave front is not traveling parallel to the dish, these distances will no longer be the same and the electric field will destructively interfere at the focus (b). 47

2.3 A sample power response profile for a 10 meter dish observing at a 1.3 millimeter wavelength. Note that most of the power is contained within λ/D radians of angular offset. This intrinsic property of parabolic reflectors leads to the fundamental relationship for single-dish resolution in Equation 2.2. 48

2.4 **Simplified Interferometry Diagram:** Light is emitted from a distant source and arrives at the telescopes as a plane wave in the direction $\hat{s}$. An additional distance of $B \cdot \hat{s}$ is necessary for the light to travel to the farther telescope, introducing a time delay between the received signals that varies depending on the source's location in the sky. The time-averaged correlation of these signals is a sinusoidal function related to the location of the source. This insight is generalized to extended emissions in the van Cittert-Zernike theorem and used to relate the time-averaged correlation to a Fourier component of the emission image in the direction $\hat{s}$. 51

2.5 The effect of taking the time-averaged correlation of two signals that are separated by a time delay. The signals received by the two telescopes are shown in red and blue, respectively. By multiplying these signals we obtain the green curve. And by averaging the values in the green curve we obtain the time-averaged correlation, shown as the straight black line. This resulting time averaged correlation is related to the phase shift between the signals, but it does not provide information about the absolute time delay. 53

2.6 The time-averaged correlation for a single point source plotted as a function of its position on the celestial sphere. Notice that as the point source moves it sweeps out a sinusoidal pattern in the time-averaged correlation. The frequency of this sinusoidal pattern is a function of the distance between the two telescopes, B, and the observing wavelength, λ. 54

2.7 **Earth Rotation Synthesis:** For every two telescopes in the interferometric array, we obtain a single measurement (visibility) related to the underlying source image's 2D spatial frequency. This frequency is related to the baseline between the telescopes in the direction perpendicular to the observing source. This would be a prohibitively small number of measurements to make an image from. However, as the Earth rotates, the projected baselines change, and we observe new measurements related to different regions of the 2D frequency plane. As time increases (specified by the Greenwich Sidereal Time (GST)), the projected baselines change and the red dots on the lower frequency coverage plot, indicating the current measurements, change position. Assuming the source is static, this amounts to carving out elliptical paths in the frequency plane that are all related to the same image. This is visualized as the transparent blue lines. As light emitted from the source is real-valued we obtain two measurements on opposite sides of the frequency plane – each independent set of measurements displayed as either the open or closed red circles. . 58

2.8 The uv coverage of Sgr A* and M87 for the same telescope array observing over 24 hours. Telescopes in the 2017 EHT array observing at a 1.3 mm wavelength were used to generate these uv-coordinates. Notice that although the telescope array has remained unchanged for these two sources, the uv-coverage changes substantially. This is due to a difference in their source direction, $\hat{s}$. 67

2.9 The observed angle of inclination affects a black hole's appearance. Here we show a prediction of what a non-spinning black hole might look like at different angles ranging from face-on (0°) to edge-on (89°). [23] . . . 70

2.10 The spin of a black hole affects its appearance. Here we show a prediction of what a black hole might look like edge-on (top) and face-on (bottom) for a variety of different spins. 0.0 spin implies that the black hole is not spinning, while 1.0 spin implies it is maximally spinning. [23] 70

3.1 **CLEAN Reconstruction Approach:** The uv-coverage of an observation characterizes the PSF of the interferometer. The PSF, referred to as the dirty beam, is found by taking the inverse Fourier transform of the uv-coverage. The dirty image, the inverse Fourier transform of the complex visibility measurements, can be described as the original image convolved with the dirty beam. To recover an image, CLEAN iteratively removes sidelobes from the dirty image caused having a non-point like dirty beam. Afterwards, the final image is blurred by the nominal beam. This process only is appropriate when there is no atmospheric error in the visibility measurements. When atmospheric error exists, it is not possible to recover the underlying image simply by removing sidelobes. In this case a process called self-calibration is employed. The uv-coverage used in this example is generated according to the specifications of 3.4.1. Images shown are 64×64 pixels and the intensity of each pixel is specified in milli-janskys. The total flux of the original source image is 1 jansky. . 78

3.2 **Image Representation:** An example of a continuous 1D image defined in terms of 1D triangle pulses. Pulses are shifted and scaled by a discrete set of values, $\boldsymbol{x}$. These shifted and scaled pulses are then summed together to make a single continuous image. In this example $\boldsymbol{x} = [1, 2, 3, 1, 1, 4]$. Note how triangle pulses result in a piecewise linear continuous image. 82

3.3 **Modeling Error for Different Continuous Image Representations:** Accurately modeling the frequencies of an image is crucial for fitting VLBI measurements during image reconstruction. Here we show, that with the same number of parameters, we can much more accurately model the true frequency distribution. A slice of frequencies for the true image is shown in red. Overlaid we show the effect of using the traditional discretized imaging model (green), and our improved model for rectangle (cyan) and triangle (blue) pulses. The dotted lines denote the frequency range sampled in the uv-coverage of Fig 3.1. Representing an image using triangle pulses reduces modeling errors for higher frequencies during image reconstruction. 84

3.4 **Intrinsic Maximum Resolution:** The geometry of a telescope array imposes an intrinsic maximum resolution on images reconstructed from its measurements. Recovering spatial frequencies higher than this resolution is equivalent to super-resolution. For results presented in this chapter, the minimum recoverable fringe spacing (corresponding to the maximum frequency) is 24.72 μ-arcseconds. The original 'Source' images (183.82 μ-arcsecond FOV) are used to synthetically generate realistic VLBI measurements. We show the effect of filtering out spatial frequencies higher than the minimum fringe spacing for these source images in 'Max Res'. 95

3.5 **Training Image Datasets:** The multivariate image prior and GMM patch prior were trained using using three distinct categories of images: natural (everyday), astronomical, and black hole simulation images. In this figure we show four random image samples from each dataset. Notice that each set of images has a very distinct look. A set of 400 natural images from the Berkeley segmentation dataset were used in the natural dataset [90]. Fifty (50) images from NASA/JPL-Caltech and NRAO were used in the astronomical dataset [2]. A set of 9000 computer generated images of Sgr A* for different inclinations and spins were used in the black hole dataset [23]. Additional images from this black hole dataset are shown in Figures 2.9 and 2.10. 97

3.6 **Trained Image Model:** Four images are sampled from each of the trained Gaussian image priors. These are shown below each model name. The top ten eigen-emission images corresponding to each model are shown directly below the image samples. Each image model captures very different characteristics from its set of images. For instance, unlike in the natural image model, the astronomical image model concentrates most of its energy in the center of the image. 99

3.7 **Trained Patch Model:** Five patches sampled from the ten most likely clusters of each GMM patch model (left to right corresponds to more to less likely). Although similar features exist in all three patch models, there are also significant differences. Each patch model appears to capture characteristics specific to its own set of images. 100

3.8 **Comparing Reconstructions from Trained Models:** Reconstructions using image priors (Gauss) and patch priors (CHIRP) trained on natural, celestial, and synthetic black hole images. The truth and blurred truth image are shown on the top. Results of Gauss are shown when constraining the complex visibilities (without atmospheric error), amplitude and bispectrum, as well as just the bispectrum. CHIRP results are only shown when constraining the bispectrum. Notice that significant differences exist between reconstructions when using the different trained image priors, even when clean complex visibilities are provided. However, only small differences can be seen in the reconstructed images using patch priors. Since absolute position is lost during imaging with the bispectrum, shifts in the reconstructed source location are expected. 101

3.9 **Comparing Reconstructions from Trained Models:** Refer to the caption of Figure 3.8. In this example the true source image comes from the manifold of images that the 'black hole' prior was trained on. However, despite this the reconstruction using 'Gauss' is very poor. This has to do with the fact that the multivariate Gaussian model is not able to adequately capture the manifold of black hole images. As all of the images used to train the 'black hole' prior are very similar, the resulting model encourages high frequency structure and negative light (see Figure 3.6). This model allows us to easily overfit the data, and find an image that aligns with the noisy data better than the true image. . 102

3.10 **Method Comparison:** Comparison of our algorithm, 'CHIRP' to three state-of-the-art methods: 'CLEAN', 'SQUEEZE', and 'BSMEM'. We show the normalized reconstruction of a variety of black hole (a-b), astronomical (c-f), and natural (g) source images with a total flux density (sum of pixel intensities) of 1 jansky and a 183.82 μ-arcsecond field of view. Since absolute position is lost when using the bispectrum, shifts in the reconstructed source location are expected. The 'TARGET' image shows the ground truth emission filtered to the maximum resolution intrinsic to this telescope array. 104

3.11 **Quantitative Analysis on Blind Test Set:** Box plots of MSE and SSIM for reconstruction methods on the blind dataset presented in Section 5.1.3. In SSIM a score of 1 implies perceptual indistinguishability between the ground truth and recovered image. Scores are calculated using the original 'Source' image (Refer to Fig. 3.4). 105

3.12 **Noise Sensitivity:** The effect of varying total flux density (in Janskys), and thus noise, on each method's recovered reconstructions. Decreasing flux results in higher noise. Notice how our method is fairly robust to the noise, while the results from other methods often vary substantially across the noise levels. The ground truth target images along with the results for a total flux density of 1 Jansky can been seen in column A and C of Figure 3.10. 106

3.13 **Real Measurements:** A comparison of our reconstructed images to [74]'s results using CLEAN self-calibration. Note that we are able to reconstruct less blurry images, and are even able to resolve 2 separate, previously unresolved, bright emissions in blazar OJ287. Measurements were taken using the VLBA telescope array. The FOV for each image is 1.5, 1, and 1 milli-arcsecond respectively. 107

4.1 **Simulated data under a static vs. varying source:** Contrasting of data observed from a static emission region (magenta) to that of a varying emission region (blue) over the course of 2.5 hours. Although both sequences start with the same image, the visibility amplitude and closure phase both begin to deviate from the static image very quickly. The ideal observation for the static and time-varying source is shown by the solid red and blue lines, respectively. We also show sample measurements with their respective error bars in the same colors. This data is simulated using the EHT2017 array from the frames in Video 3 presented in Section 4.4. 111

4.2 **Graphical Representation of our Dynamic Imaging Model**: At each time t we observe a vector of data products $\boldsymbol{y}_t$ corresponding to the instantaneous source image $\boldsymbol{x}_t$. We assume each image $\boldsymbol{x}_t$ is related to its adjacent neighbors in time, $\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_{t+1}$, and is also related to a multivariate Gaussian distribution specified by mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Lambda}$. The persistent global evolution of the source images over time is specified by $\boldsymbol{A}$, which is further parameterized by θ. Additional intensity perturbations in time are captured by the covariance matrix $\boldsymbol{Q}$. In this diagram, squares indicate parameters, circles are variables, and shaded circles indicate the variable is observed. 113

4.3 **Static Imaging Comparison:** Results of static imaging using a multivariate Gaussian prior (a = 2, 5, 10) compared to state-of-the-art reconstruction methods using MEM & TV regularizers [26] as well as patch-based regularizers (CHIRP) (see Chapter 3 and [17]). All images are shown with a field of view of 160 μ-arcseconds. Data is generated using a static image with the uv-coverage of the EHT2017 array shown on the left (see Section 4.4). The uv-coverage is colored by time, as indicated by the colorbar in Figure 4.6. Although the previous algorithms (MEM & TV and CHIRP) both produce better results, the Gaussian reconstruction is able to correctly get the broad structure of the underlying image. Since we do not impose positivity, negative values are reconstructed. However, by clipping the resulting image we can see that the result aligns well with the true static image. The Gaussian prior model also allows us to easily estimate our reconstructed image uncertainty. We visualize the diagonal entries of the posterior covariance matrix as the reshaped standard deviation image. Note that as the smoothness parameter a is increased, the per-pixel standard deviation becomes smaller, but the structure of the standard deviation deviates from what was specified in the prior (recall $\mathbf{\Lambda}$ is scaled by $\boldsymbol{\mu}$, which we have specified as a 2D Gaussian in this work). For large a the uncertainty is shown to be primarily in the diagonal north-west to south-east direction, due to the lack of spatial frequencies sampled by the telescope array in this direction. To avoid approximations and best show the recovered posterior covariance matrices, atmospheric error has not been included in the data used to recover these images. The flux in Janskys per pixel is normalized across images for a 30 $\times$ 30 image. Note we have flipped the uv-coverage relative to how it is normally shown to align with regions that have missing frequency information in the image (in the rest of this thesis we show the uv-coverage plotted as if we were looking towards the Earth rather than towards the source). 115

4.4 **Gaussian Image Prior:** The covariance matrix constructed for $a = 2, 3, 4$ along with image samples from the prior $\mathcal{N}_x(\mu, \Lambda)$. The image samples have a field of view of 160 μ-arcseconds. Notice that as a increases, the sampled images appear smoother. In these examples μ is a 2D Gaussian Image with standard deviation of 75 μ-arcseconds. and $b = 0.5$. Note that as a increases, the prior encourages smoother structure. 116

4.5 **Ground truth videos:** The four ground truth sequences used to demonstrate results. We show a single frame from each sequence, the mean frame, and the spatial standard deviation of flux density. Video 1 consists of a 160μ-arcsecond image [23] that rotates 180° over the course of a 12 hour observation (24 hour rotational period). Video 2 is a 120μ-arcsecond view of an edge-on black hole disk with a rotating "hot spot" predicted by [21] with a rotational period of 2.78 hours. Video 3 and 4 are generated using a GRMHD model of a black hole observed face on and at a 45° inclination with a 160μ-arcsecond field of view [118]. They assume a spin of 0.9375 with an Innermost Stable Circular Orbit (ISCO) rotational period of 8.96 minutes. 122

4.6 **Time-varying uv-coverage:** The uv-coverage for EHT2017, EHT2017+ and FUTURE arrays when observing SgrA. Baselines are colored by the time of each observation relative to the start time, indicated by the colorbar to the right. 123

4.7 **Static evolution model:** Results obtained using data simulated from each of the 4 video sequences (see Figure 4.5) under different telescope arrays (see Figure 4.6) and noise conditions. The main portion of the figure is broken up into 4 quadrants corresponding to Videos 1-4 when moving from left to right, top to bottom. The true mean image from the ground truth videos, blurred to 3/4 the nominal resolution of the array, is shown on the top. We compare results of our proposed method, StarWarps, to that of the single imaging methods presented in [87] and [26]. In particular, we compare the mean image obtained using StarWarps video reconstruction. The error type NO ATM. indicates reconstructing using visibilities on data with no atmospheric error, while the error type ATM. indicates using the visibility amplitudes and bispectrum on data where atmospheric phase errors have been introduced. The quality of each result, compared to the ground truth mean image, is indicated in the table of normalized root mean squared errors (Normalized RMSE). To account for the loss of absolute position in the presence of atmospheric phase error, images were rigidly aligned to the true mean before computing the error. 125

4.8 **Recovering Warp Field:** By solving for the parameters of a persistent warp field using the proposed EM algorithm, we are able to recover a low-dimensional representation of the source dynamics. Results are shown using the EHT2017+ array with and without atmospheric error (ATMOSPHERIC and NO ATMOSPHERIC ERROR, respectively). Arrows showing the direction of recovered motion are overlaid on the mean image for a recovered video. In Video 1 the true underlying motion can be described by a clockwise rotation. The proposed method is able to recover Video 1's motion from the observed data. Video 2 contains a 'hot spot' rotating counter-clockwise around a static emission. Video 2 cannot be described using a single persistent flow field. Yet, despite this, the proposed method is still able to recover the general direction of counter-clockwise motion. 127

4.9 **Time-resolved reconstruction of Video 1:** Video 1 contains an image rotating clockwise by 180° over the course of the observation. At each time, the interferometric telescope array measures values related to 2D spatial frequencies of the current underlying image, shown in the row labeled 'Truth'. These are indicated by the dots on the uv-coverage plots (each independent set of measurements displayed as either black or red). We present results obtained when using calibrated data with no atmospheric error, as well as when there is atmospheric phase error still present and we must use data products invariant to its effects. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained when we assume a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. In the case that we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image. 128

4.10 **Time-resolved reconstruction of Video 2:** Video 2 contains a sequence of a hotspot orbiting counter-clockwise around a black hole. We present time-resolved results obtained using data derived from this sequence. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained assuming a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. If we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image. Our method substantially improve results over the snapshot method, especially in the case of atmospheric error when the absolute position of the source cannot be recovered. Additionally, despite the fact that this hotspot video does not match our assumed motion model, using our proposed approach we were able to estimate a warp field that provides the direction of the source's true underlying motion. See the caption of Figure 4.9 for more detail. . 129

4.11 **Visualizing Recovered Motion:** We visualize the recovered motion in Video 2 by displaying the change in intensity around a circle in the image over time. After fitting a circle of constant radius to each video, the intensities around the circle in each image are unwrapped and placed in a single column in the unwrapped space × time image. As the hot spot rotates around the black hole a distinctive line appears in the true angle × time image. These lines also appear in the StarWarps angle × time images, but are harder to discern among the other artifacts in the snapshot imaging result. Results were obtained using the EHT2017+ array with added atmospheric noise, and correspond to results shown in Figure 4.10. As the absolute position of the source is lost when using the closure phase or bispectrum, the position of the recovered black hole moves slightly over the course of the video. This causes the fluctuation in the intensity of the bright horizontal line in the StarWarps recovered angle × time images, as we do not shift the position of the fitted circle. 131

5.1 The EHT Imaging Challenge Pipeline: The imaging challenges consist of four steps: simulation, reconstruction, evaluation, and learning. This process encourages algorithm development and improves our understanding of imaging methods on EHT quality data. This newly acquired knowledge then helps to shape the next imaging challenge. 137

5.2 A screen shot of the evaluation website used by judges in the fourth challenge (dataset 7). Judges were provided with the submission images, challenge data, and plots that showed how well each image fit the simulated measurements. Additional plots (such as the uv-coverage, beam size, etc) were provided upon request. 142

5.3 Dataset 1: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 144

5.4 Dataset 1: Truth (top) and submitted images in two colormaps. 145

5.5 Dataset 1: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 145

5.6 Dataset 2: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that amplitude mis-calibration errors cause the amplitude to sometimes exceed the emission's total flux of 2 Jy. 147
5.7 Dataset 2: Truth (top) and submitted images in two colormaps. 148
5.8 Dataset 2: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 148
5.9 Dataset 3: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 150
5.10 Dataset 3: Truth (top) and submitted images in two colormaps. 151
5.11 Dataset 3: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 151
5.12 Dataset 4: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 153
5.13 Dataset 4: Truth (top) and submitted images in two colormaps. 154
5.14 Dataset 4: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 154
5.15 Dataset 5: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 156
5.16 Dataset 5: Truth (top) and submitted images in two colormaps. 157
5.17 Dataset 5: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 157
5.18 Dataset 6: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that the amplitude miscalibration errors causes the amplitude to sometimes exceed the emission's total flux of 2 Jy. 159
5.19 Dataset 6: Truth (top) and submitted images in two colormaps. 160
5.20 Dataset 6: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 161

5.21 Dataset 7: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 163
5.22 Dataset 7: Truth (top) and submitted images in two colormaps. 164
5.23 Dataset 7: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 165
5.24 Dataset 8: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that amplitude mis-calibration errors cause the amplitude to sometimes exceed the emission's total flux of 2 Jy. 167
5.25 Dataset 8: Truth (top) and submitted images in two colormaps. 168
5.26 Dataset 8: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 169
5.27 Dataset 9: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 171
5.28 Dataset 9: Average truth image (top) and submitted images in two colormaps. 172
5.29 Dataset 9: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 172
5.30 Dataset 9: Frames over time from the truth video. Note that there are significant structural changes in the source image over time. 173
5.31 Dataset 10: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. 175
5.32 Dataset 10: Average truth (top) and submitted images in two colormaps. 176
5.33 Dataset 10: Frames over time from the truth video. Note that there is substantial time-variability, but it the variation is significantly less than in Dataset 9 (see Figure 5.30). 177
5.34 Dataset 10: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images. 177

5.35 Images submitted to the first four imaging challenges by Akiyama & Tazaki using Sparse Modeling. The truth image is shown on the top and the submitted image right below. Note the substantial improvement of their method in recovering the correct structure of a static source. . . . 179
5.36 For extended emissions that require a large field of view during imaging, the imaging challenges helped to identify that it was very important to constrain the visibility amplitudes. As the bispectrum is the product of three terms, there can be a lot of uncertainty in reconstructing each visibility's amplitude when imaging purely with the bispectrum. This can result in poor reconstructions, especially when the field of view is not well constrained. 181
5.37 A case of degeneracy in EHT data. Two possible explanations fit the simulated EHT data equally well. In most cases the closure phases appeared nearly identical, and only small differences existed (that fit well within the error expected on measurements). We believe this degeneracy exists due to small closure phases with large error. 182
5.38 The black hole's lensed photon ring is highlighted by the yellow ring. However, a funnel wall causes a more prominent ring feature that shows up in the reconstruction, highlighted in red. This example emphasizes the importance of further study into features that may masquerade as a black hole photon ring. 183
5.39 Five submissions reconstructed using the same data but using different imaging algorithms. 184

6.1 We construct a 1-D video of an obscured scene using RGB video taken with a consumer camera. The stylized diagram in (a) shows a typical scenario: two people—one wearing red and the other blue—are hidden from the camera's view by a wall. Only the region shaded in yellow is visible to the camera. To an observer walking around the occluding edge (along the magenta arrow), light from different parts of the hidden scene becomes visible at different angles (see sequence (b)). Ultimately, this scene information is captured in the intensity and color of light reflected from the corresponding patch of ground near the corner. Although these subtle irradiance variations are invisible to the naked eye (c), they can be extracted and interpreted from a camera position from which the entire obscured scene is hidden from view. Image (d) visualizes these subtle variations in the highlighted corner region. We use temporal frames of these radiance variations on the ground to construct a 1-D video of motion evolution in the hidden scene. Specifically, (e) shows the trajectories over time that specify the angular position of hidden red and blue subjects illuminated by a diffuse light. 186

6.2 In (a), the transfer matrix, $\mathbf{A}$, is shown for a toy situation in which observations lie along circles around the edge. In this case, $\mathbf{A}$ would simply be a repeated lower triangular matrix. (b) contains an example estimation gain image, which describes the matrix operation performed on observations $\mathbf{y}^{(t)}$ to estimate $\mathbf{x}^{(t)}$. As predicted, the image indicates that we are essentially performing an angular derivative in recovering a frame of the 1-D video. 189

6.3 One-dimensional reconstructed videos of indoor, hidden scenes. Results are shown as space-time images for sequences where one or two people were walking behind the corner. In these reconstructions, the angular position of a person, as well as the number of people, can be clearly identified. Bright vertical line artifacts are caused by additional shadows appearing on the penumbra. We believe horizontal line artifacts result from sampling on a square grid. 192

6.4 1-D reconstructed videos of a common outdoor, hidden scene under various weather conditions. Results are shown as space-time images. The last row shows results from sequences taken while it was beginning to rain. Although artifacts appear due to the appearing raindrops, motion trajectories can be identified in all reconstructions. 192

6.5 The result of using different cameras on the reconstruction of the same sequence in an indoor setting. Three different 8-bit cameras (an iPhone 5s, a Sony Alpha 7s, and an uncompressed RGB Point Grey) simultaneously recorded the carpeted floor. Each camera introduced a different level of sensor noise. The estimated standard deviation of per-pixel sensor noise, λ, is shown in (b). We compare the quality of two sequences in (c) and (d). In (c), we have reconstructed a video from a sequence of a single person walking directly away from the corner from 2 to 16 feet at a 45 degree angle from the occluded wall. This experiment helps to illustrate how signal strength varies with distance from the corner. In (d), we have done a reconstruction of a single person walking in a random pattern. In (c) the hidden person does not change in angular position. Thus, for these results, we subtract an average background frame computed from a different portion of the video sequence. 193

6.6 A subject's reconstructed angular velocity relative to the corner as a function of time. In this sequence, a person was walking in circles far from the corner. 195

6.7 The four edges of a doorway contain penumbras that can be used to reconstruct a 180° view of a hidden scene. The top diagram indicates the penumbras and the corresponding region they describe. Parallax occurs in the reconstructions from the left and right wall. This can be seen in the bottom reconstruction of two people hidden behind a doorway. Numbers/colors indicate the penumbras used for each 90° space-time image. 196

6.8 A hidden person will introduce an intensity change on the left and right wall penumbras at angles of $\theta_L^{(t)}$ and $\theta_R^{(t)}$, respectively. Once these angles have been identified, we can recover the hidden person's two-dimensional location using Eq. 6.11. 197

6.9 The results of our stereo experiments in a natural setting. Each sequence consists of a single person walking in a roughly circular pattern behind a doorway. The 2-D inferred locations over time are shown as a line from blue to red. Error bars indicating one standard deviation of error have been drawn around a subset of the points. Our inferred depths capture the hidden subject's cyclic motion, but are currently subject to large error. A subset of B's inferred 2-D locations have been cut out of this figure. 199

6.10 Controlled experiments were performed to demonstrate the ability to infer depth from stereo edge cameras. A monitor displaying a moving green line was placed behind an artificial doorway (a) at four locations corresponding to 23, 40, 60, and 84 cm, respectively. (b) shows sample reconstructions done of the edge cameras for the left and right wall when the monitor was placed at 23 and 84 cm. Using tracks obtained from these reconstructions, the 2-D position of the green line in each sequence was estimated over time (c). The inferred position is plotted with empirically computed error ellipses (indicating one standard deviation of noise). 200

6.11 The empirical means plus or minus one standard deviation of the estimated P_z as a function of its x-coordinate, assuming true P_z of 20, 40, 60, and 80. Here, the two corner location errors at each of the boundaries of the doorway are independent and subject to $\sigma^2_{\Delta x} = \sigma^2_{\Delta z} = 0.04$. . . 200

C.1 This figure shows the setup for the toy problem of interest. The scene consists of a single bright object, whose angular position θ we want to learn. 222

C.2 This figure shows the impact of a corner location error. In the error-free case, we would sweep ϕ across the observation plane (shown in green) hinging around the corner (the solid black line). But if made a corner location error, we would instead try to sweep ϕ across the observation plane erroneously (shown in blue) hinging around the false corner (shown with a dotted line). 223

C.3 This plot shows how the observed intensity values vary with ϕ, in the case of correct corner location (in blue) and a corner location error ($(d_x, d_y) = (0.1, 0.2)$, in red). Note that in the case of a corner location error, the maximum value of the intensity does not reach 1. Note also that d_x and d_y are as a fraction of the radius of the observation plane, r, which here is taken to be 1. 224

C.4 This plot is intended as a reference for the meanings of each of the variables used in the calculations of ϕ as a function of θ. 224

C.5 This plot shows the empirical mean (in blue) plus or minus one standard deviation (in red) of the error as a function of θ. Here, $\sigma_x = 10^{-4}$ and $\sigma_y = 10^{-3}$. 225

C.6 The empirical means plus or minus one standard deviation of the estimated P_z as a function of its x-coordinate, assuming true P_z of 20, 40, 60, and 80. Here, the two corner location errors at each of the boundaries of the doorway are independent and subject to $\sigma^2_{\Delta x} = \sigma^2_{\Delta z} = 0.04$. We sample from a set of 1000 corner errors to approximate the mean and standard deviations empirically. 226

C.7 The reconstructed depths of objects at depths 1, 2, 3, and 4, given a corner error of $\Delta y_1 = \Delta y_2 = 0.02$. 227

C.8 The reconstructed depths of objects at depths 1, 2, 3, and 4, given a corner error of $\Delta y_1 = -0.02$, $\Delta y_2 = 0.02$. Note that because of the different corner errors for each corner, there is the possibility of asymmetric behavior on either side of the doorway. 227

C.9 The result of imposing temporal smoothness or averaging adjacent frames in time to help in reducing noise. 228

1

Introduction

WATCHING a raindrop fall into a puddle of water, causing a ring of expanding ripples, may seem mundane. However, behind this commonplace phenomenon is a complex process governed by fluid dynamics and its interaction with light. By understanding these connections we realize that the water's temperature is encoded in subtle speed differences of the expanding ripples and the size of the falling water droplet [12]. With the proper models and analysis techniques, we may be able to extract this seemingly invisible signal and turn a regular camera into a visual thermometer, without ever interacting with the fluid.

As science continues to push the boundaries of knowledge, sensors are reaching the limits of what can be measured using traditional-direct approaches [11,58,59]. Thus we must think of creative, new strategies to recover information from indirect measurements. The simple raindrop example above illustrates how looking at the world from a unique, and often non-obvious, perspective can make seemingly impossible problems tractable. By incorporating domain-specific knowledge, and exploiting structure in data, we can often find new ways to indirectly recover the desired information.

Imaging often plays a critical role in advancing fundamental science [11,20,63]. Through a combination of novel processing and sensing strategies, it is possible to design imaging systems that exceed fundamental theoretical limitations to see deeper, farther, and faster than ever before [11]. However, these non-traditional imaging systems generally come with a trade-off; researchers are left with increasingly sparse and/or noisy measurements that require incorporating additional structure to extract anything meaningful. The focus of this thesis is on using computational methods that exploit structure in our universe to uncover things thought to be invisible.

Many problems we have studied exploit structure in our universe to pull out information that may at first seem unobtainable. For example, by analyzing tiny motions in a seemingly still scene, and relating this motion to vibration theory, we showed in [33, 34] how it is possible to infer information about material properties. Moreover, by finding structure across a population in low-resolution clinical MRI scans, in [31] we were able to impute missing data. This thesis focuses on two imaging problems we have studied that explicitly leverage structure in our universe: reconstructing images and video from a computational telescope the size of the Earth, and seeing around corners. In Sections 1.0.1 and 1.0.2 we give a brief overview of these indirect imaging problems, and the contributions we have made to them. For the remainder of the chapter, in Section 1.1, we outline the common philosophy taken in solving these problems.

1.0.1 Imaging with an Earth-Sized Computational Telescope

It is believed that the heart of the Milky Way hosts a four million solar mass black hole feeding off a spinning disk of hot gas [50]. An image of the shadow cast by the event horizon of the black hole could help to address a number of important scientific questions [36]; for instance, does Einstein's theory of general relativity hold in extreme conditions? Unfortunately, the event horizon of this black hole appears so small in the sky that imaging it would require a single-dish radio telescope the size of the Earth. Although a single-dish telescope this large is unrealizable, by connecting disjoint radio telescopes located around the globe, the "Event Horizon Telescope" (EHT) is creating an Earth-sized computational telescope [35]. Through the use of a technique referred to as Very Long Baseline Interferometry (VLBI), it is possible to combine measurements from across this global array of telescopes to take the very first picture of a black hole's shadow [124] (see Chapter 2).

Although measurements from the EHT's computational telescope share similarities with other well-studied problems [20, 89, 95, 123], the EHT faces a number of unique challenges. To achieve the necessary resolution, the EHT must operate at very short-wavelengths and use fewer telescopes than normally employed in VLBI setups. This results in measurements that are incredibly sparse and noisy, causing traditional astronomical imaging methods to often perform poorly [17, 85]. By combining techniques from both astronomy and computer science, we have been able to develop innovative ways to robustly reconstruct the underlying images. Although these methods were specifically developed to tackle the EHT's unique challenges, they are broadly applicable to interferometric radio imaging.

In Chapter 3 this thesis introduces methods developed for single image reconstruction when using the EHT's sparse and heterogeneous array. Specifically, we discuss how to handle propagation delays caused by the atmosphere that plague EHT data, as well as introduce data-driven priors into VLBI imaging as a way to help tackle uncertainty in bias. These algorithms, like all prior VLBI imaging techniques, require that over the course of a night's observation the source remains static. Thus, in Chapter 4 this thesis addresses the violation made by our galactic center's quickly changing black hole [72, 87]. In this chapter we present an algorithm developed to reconstruct a video of the time-evolution of the targeted source over the course of a night. This allows us to not only study the structure of the emission in the immediate region around a black hole, but also its dynamics.

Current literature in astronomical imaging lacks rigor in method evaluation [10, 13, 83]. In Chapter 5 this thesis introduces efforts we have taken to improve the evaluation process. In particular, we introduce a dataset containing realistic measurements along with a website testbed. Furthermore, we discuss the EHT Imaging Challenges - an ongoing contest developed to evaluate imaging methods' ability to handle the challenges faced in EHT quality data. This thesis presents results from the first four challenges, and discusses a number of insights made. These newly introduced evaluation approaches facilitate rigorous comparisons between imaging algorithms, and have already proven to promote growth in both new and old methods.

■ 1.0.2 Imaging Hidden Scenes Behind Corners

The ability to see around obstructions would prove valuable in a wide range of applications. As just two examples, remotely sensing occupants in a room would be valuable in search and rescue operations, and the ability to detect hidden, oncoming vehicles and/or pedestrians would be valuable in collision avoidance systems [16]. Although often not visible to the naked eye, in many environments, light from obscured portions of a scene is scattered over many of the observable surfaces.

In Chapter 6 this thesis demonstrates an instance of this phenomenon, whereby we can interpret the light falling on the ground near a naturally occurring corner as a "camera" to see the scene beyond it. In particular, we demonstrate methods for using the subtle spatio-temporal radiance variations that arise on the ground at the base of edges to construct a one-dimensional video of the hidden scene. Since vertical edges are ubiquitous, such cameras can be found in many environments. From standard RGB video recordings of the variations in intensity, we use edge cameras to recover a 1-D video that reveals the number and trajectories of people moving in an occluded scene. We further show that adjacent vertical edges, such as those that arise in the case of an open doorway, yield a stereo camera from which the 2-D location of hidden, moving objects can be recovered. We demonstrate the technique in a number of indoor and outdoor environments involving varied surfaces and illumination conditions.

1.1 Inverse Problems

The goal of a sensing system is to capture and provide measurements, $\boldsymbol{y}$, that describe a desired signal, $\boldsymbol{x}$. In the case of direct sensing systems, the provided measurements tell us exactly what we are looking for. For instance, an outdoor scene can be directly captured using a pinhole camera. Similarly, the temperature of the puddle of water in the beginning of this chapter could be measured with a mercury thermometer. In both of these cases, $\boldsymbol{y}$ is roughly equivalent to $\boldsymbol{x}$. However, many desired signals cannot be directly observed. For instance, we can not use the pinhole camera to see the organs inside of a person, and perhaps we cannot stick a thermometer into the water for fear of contamination. For these tasks, since we cannot measure the signal directly, we must instead take measurements that merely depend on the unknown, desired signal (e.g X-ray absorbance through the body, or expanding ripples). In other words, the measurements are likely to be a complex, possibly non-deterministic, function of $\boldsymbol{x}$: $\boldsymbol{y} \approx f(\boldsymbol{x})$. The objective is then to recover the desired signal, $\boldsymbol{x}$, from the indirect measurements, $\boldsymbol{y}$. This process is referred to as solving the *inverse problem* [11].

An initial way you may think to approach this inverse problem may be to construct an inverse function, $f^{-1}(\boldsymbol{y})$, of $f(\boldsymbol{x})$ that acts on the measurements and results in the desired signal (see Figure 1.1). Although this is the standard way of solving many problems [57, 60, 96], this approach has significant flaws. First, the measurements, $\boldsymbol{y}$, may be related to the desired signal, $\boldsymbol{x}$, through a very complicated non-linear process. This often makes it prohibitively difficult to construct an inverse function. Second, in order for an inverse to even exist $f(\boldsymbol{x})$ must be one-to-one, meaning it cannot map different signals to the same measurements. Third, the addition of non-deterministic noise cannot be incorporated into an inverse and thus cannot be properly accounted for. Thus, when taking this approach there is often significant ad-hoc post-processing done in an attempt to correct the recovered signal [121].

Instead, a more successful approach, that is adopted in this thesis, is to construct a function of the forward system $f(\boldsymbol{x})$, and then search for the signal (or set of signals) $\boldsymbol{x}$ that would produce the observed measurements $\boldsymbol{y}$. This approach alleviates the need to construct an impossible inverse function, and instead focuses our energy on understanding and correctly modeling the sensing system. However, note that this

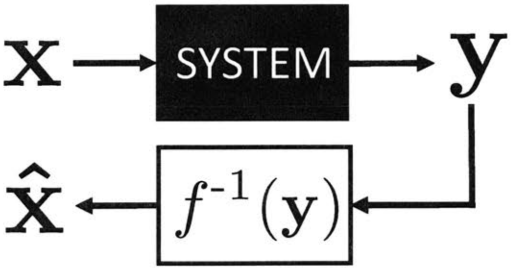


Figure 1.1. A traditional approach to solving an inverse problem might look something like what is illustrated in this diagram. The sensing system produces measurements $\boldsymbol{y}$ in response to the unknown signal $\boldsymbol{x}$. To recover the unknown signal an inverse function is constructed, $f^{-1}(\boldsymbol{y})$, which acts on the observed measurements, $\boldsymbol{y}$, to produce an estimate of the unknown signal, $\hat{\boldsymbol{x}}$. Although this approach seems fairly straightforward, it has a number of flaws and often results in poor reconstructions.

approach does not resolve how to deal with noise or non-unique system functions (eg. $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2)$). To handle this, we incorporate a prior model to characterize the likely behavior of the unknown signal. Using these two ingredients, the forward and prior model, we can define a standard recipe for obtaining the desired unknown signal (see Figure 1.2).

Without loss of generality, for the remainder of this chapter, we focus on describing how to model and infer images using indirect imaging systems. In this case, $\boldsymbol{x}$ is an image and $\boldsymbol{y}$ are the observed measurements (or data) from the imaging system. The forward system is not necessarily invertible, and thus does not necessarily have to return a unique $\boldsymbol{y}$ for each unique $\boldsymbol{x}$.

Forward Model

The forward model of an imaging system describes how an image, $\boldsymbol{x}$ is transformed and distorted into the observed measurements $\boldsymbol{y}$. This model should fully characterize the system by including all deterministic (e.g. geometry) and non-deterministic properties (e.g. noise). The forward model is embodied in the probability, $p(\boldsymbol{y}|\boldsymbol{x})$, which describes the distribution of measurements, $\boldsymbol{y}$, conditioned on $\boldsymbol{x}$ being the true underlying image.

The conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ can be generically defined. However, it is often the case that systems deal with noise that is due to independent random perturbations. For instance, thermal noise due to agitation of the electrons in a circuit can be described in this manner, and affects both imaging systems studied in this thesis: telescope receivers as well as digital cameras [51]. In these cases, the central limit theorem explains

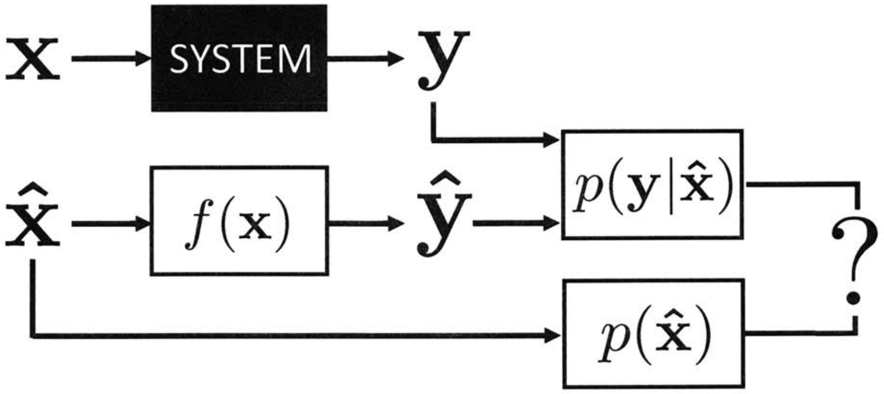


Figure 1.2. In this thesis a Bayesian style inference approach is taken to solve inverse problems. The sensing system produces measurements $\boldsymbol{y}$ in response to the unknown signal $\boldsymbol{x}$. To recover the unknown signal we construct a function$f(\boldsymbol{x})$ which models the behavior of the system. We then search for a signal $\hat{\boldsymbol{x}}$ that would produce measurements, $\hat{\boldsymbol{y}}$, similar to what we have observed and also fits our model of likely behavior of the unknown signal, $\boldsymbol{x}$.

why this type of noise follows a zero-mean Gaussian distribution [49]. Mathematically,

$$\boldsymbol{y} \sim \mathcal{N}(f(\boldsymbol{x}), \boldsymbol{\Sigma}), \tag{1.1}$$

where $\mathcal{N}_z(m, \Sigma)$ is the multivariate normal distribution of z with mean m and covariance Σ. The function $f(\boldsymbol{x})$ embodies all of the deterministic properties of the system, and $\boldsymbol{\Sigma}$ describes the covariance of system noise.

Domain-Specific Knowledge Making substantial advances in imaging requires the ability to bridge the gap between theory and practice. Thus, domain-specific knowledge is invaluable when defining an imaging system's forward model. Understanding the minutiae of a specific system, and properly accounting for its differences compared to generic systems, is often crucial in recovering the best possible image, $\boldsymbol{x}$.

A Gaussian distribution is often the de facto assumption made about noise on imaging systems. However, there are many types of noise/error that can affect a system that cannot be described in this manner [32]. For instance, Poisson/shot noise appears in digital cameras in low-light situations [88]. This noise cannot be modeled with a Gaussian distribution, and must be accounted for in particular situations to avoid serious reconstruction artifacts. Likewise, propagation delays due to the atmosphere plague measurements for the Event Horizon Telescope and also cannot be modeled as Gaussian noise [124]. If this error is not accounted for during imaging, resulting image reconstructions are meaningless.

We have put a large emphasis on collaborating closely with experts in VLBI for developing EHT imaging techniques. These experiences, including fieldwork at an EHT participating telescope, have made it possible for us to better understand the system we are working with, and adapt imaging methods to account for its unique properties.

Prior

Using just a forward model, it is possible to search for an image $\hat{\boldsymbol{x}}$ that maximizes $p(\boldsymbol{y}|\boldsymbol{x})$. However, due to noise on $\boldsymbol{y}$ and the large dimensionality of parameters (pixels) in $\boldsymbol{x}$, the result of this search will generally find an $\hat{\boldsymbol{x}}$ that looks nothing like the original image that generated the observed measurements, $\boldsymbol{y}$. The resulting $\hat{\boldsymbol{x}}$ will often contain large amounts of noise and artifacts. It may even violate fundamental physical constraints, such as generate an image with negative light.

The reason simply maximizing the conditional probability results in a poor reconstruction is due to the fact that the inherent likelihood of each image is not being accounted for. In fact, by only maximizing $p(\boldsymbol{y}|\boldsymbol{x})$ we are assuming all possible combinations of pixels in $\boldsymbol{x}$ are equally likely. However, as our universe contains a lot of structure, this is clearly a very poor assumption; an image where each pixel is independently sampled is incredibly unlikely to produce an image of something we are likely to observe.

When solving for $\hat{\boldsymbol{x}}$ we instead should incorporate models that try to characterize this structure by approximating the distribution $p(\boldsymbol{x})$ of images in our universe (or the specific domain in which we are working). Image models can often be described using many fewer parameters than there are pixels in the image [122]. Proof of this intrinsic image property is in JPEG compression, which reduces the number of parameters used to describe an image with almost no perceptual loss [54]. Modeling the true distribution of possible images makes it possible to incorporate our prior knowledge of what we expect images to look like into our inverse imaging system.

■ 1.1.1 Bayesian Inference

After defining the forward and prior model, the task of combining information and finding the optimal image, $\boldsymbol{x}$, given the measurements, $\boldsymbol{y}$, still remains a challenge. From Bayes' theorem [14] we can write the probability of $\boldsymbol{x}$ being the true underlying image that produced measurements $\boldsymbol{y}$ as:

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &= \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} && (1.2)\\
&\propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}). && (1.3)
\end{aligned}
$$

Using the forward model we can define a conditional distribution of the measurements, $\boldsymbol{y}$, given the unknown image, $\boldsymbol{x}$: $p(\boldsymbol{y}|\boldsymbol{x})$. Similarly, the prior model allows us to define $p(\boldsymbol{x})$. Once we have defined each of these terms, finding the optimal image, $\hat{\boldsymbol{x}}$, then simply reduces to maximizing the posterior distribution. Mathematically,

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left[p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \right]. \tag{1.4}$$

As log is a monotonic function, maximizing the posterior distribution is equivalent to minimizing the negative log-likelihood [14]:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[-\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{x}) \right]. \tag{1.5}$$

Note that Equation 1.5 has the simple form adopted by many regularized inverse problems:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[\chi(\boldsymbol{x}, \boldsymbol{y}) - \beta \mathcal{R}(\boldsymbol{x}) \right], \tag{1.6}$$

where $\chi(\boldsymbol{x}, \boldsymbol{y})$ indicates how inconsistent the image, $\boldsymbol{x}$, is with the observed data, $\boldsymbol{y}$, and $\mathcal{R}(\boldsymbol{x})$ expresses how likely we are to have observed the image $\boldsymbol{x}$. These two terms often have different preferences for the "best" image, and fight against each other in selecting $\hat{\boldsymbol{x}}$. Their relative power in this decision is specified with the hyper-parameter β. Equation 1.6 can be interpreted probabilistically when $-\chi(\boldsymbol{x}, \boldsymbol{y}) = logp(\boldsymbol{y}|\boldsymbol{x})$, $\mathcal{R}(\boldsymbol{x}) = \log p(\boldsymbol{x})$, and $\beta = 1$. While many methods do not have a probabilistic interpretation, their formulation leads to a similar optimization as a probabilistically motivated model.

Gaussian Distributions

For general distributions, Equation 1.5 can be solved using generic methods such as gradient descent. However, in the case of certain special distributions, we may be able to obtain a deeper understanding of this optimization. A common approximation is to define $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$ in terms of multivariate Gaussian distributions [14]. Mathematically,

$$\boldsymbol{y} \sim \mathcal{N}(f(\boldsymbol{x}), \boldsymbol{\Sigma}) \tag{1.7}$$
$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{1.8}$$

In this case, the distributions of $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$ are written as:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(f(\boldsymbol{x}) - \boldsymbol{y})^T\boldsymbol{\Sigma}^{-1}(f(\boldsymbol{x}) - \boldsymbol{y})\right] \tag{1.9}$$
$$p(\boldsymbol{x}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Lambda}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]. \tag{1.10}$$

Under these Gaussian models, Equation 1.5 reduces to a simple expression:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[\left[(f(\boldsymbol{x}) - \boldsymbol{y})^T\boldsymbol{\Sigma}^{-1}(f(\boldsymbol{x}) - \boldsymbol{y})\right] + \left[(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]\right]. \tag{1.11}$$

For generic $f(\boldsymbol{x})$ this is still difficult to solve. However, in the case that $\boldsymbol{y}$ is a linear function of $\boldsymbol{x}$, a closed-form solution of $\hat{\boldsymbol{x}}$ can be found [14]. Let, $f(\boldsymbol{x}) = \boldsymbol{F}\boldsymbol{x}$. Then, we can compute the most likely estimate of each $\boldsymbol{x}$ as:

$$\hat{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F}^T(\boldsymbol{\Sigma} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}(y - \boldsymbol{F}\boldsymbol{\mu}). \tag{1.12}$$

This is often referred to as Wiener filtering [106, 132]. Note that in the limit of having no prior information about the underlying image $\boldsymbol{x}$, e.g $\boldsymbol{\Lambda} = \lim_{\lambda\to\infty} \lambda\mathbb{1}$, this MAP solution reduces to $\hat{\boldsymbol{x}} = \boldsymbol{F}^T(\boldsymbol{F}\boldsymbol{F}^T)^{-1}\boldsymbol{y} = \boldsymbol{F}^{-1}\boldsymbol{y}$. Thus, in the worse case scenario, this solution reduces to finding and applying an inverse function. Note this is only possible if an inverse exists ($\boldsymbol{F}$ is full rank).

Equation 1.12 can be found through multiple different methods. The most natural way to solve Equation 1.11 is to solve for where the derivative of the negative log-likelihood equals zero. However, an alternative approach is to evaluate the full posterior distribution. Since the product of Gaussian densities is Gaussian, for $f(\boldsymbol{x}) = \boldsymbol{F}\boldsymbol{x}$,

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}_{\boldsymbol{x}}(\hat{\boldsymbol{x}}, \boldsymbol{C}). \tag{1.13}$$

The mean (and mode) of this posterior distribution is equivalent to Equation 1.12. However, this distribution not only tells us the mean (and mode), but also the uncertainty in the estimate through the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{F}^T(\boldsymbol{\Sigma} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{\Lambda}. \tag{1.14}$$

Estimating this uncertainty is useful in understanding what regions of the reconstructed image we trust, and can become especially valuable when propagating information across disjoint observations. Refer to Appendix A.1 for a derivation of this result.

1.1.2 Bias-Variance Trade-off

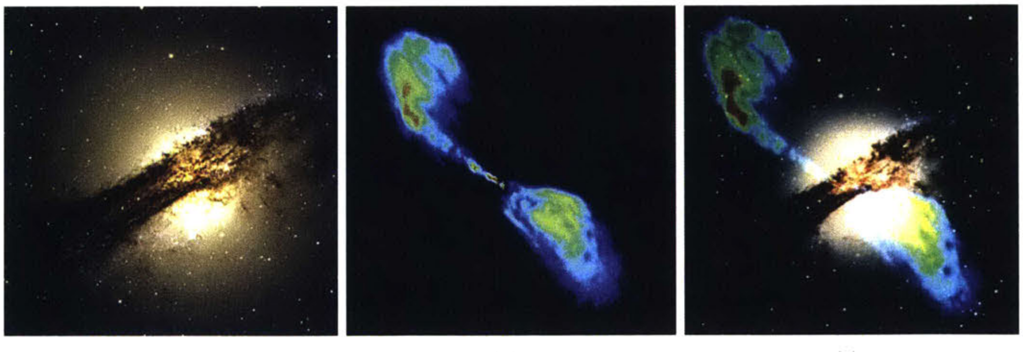
One disadvantage of Bayesian imaging is that bias is often introduced in the recovered image, $\hat{\boldsymbol{x}}$. As the prior model is unlikely to perfectly match the true distribution of images, it will slightly push the solution in a direction it favors, potentially misrepresenting the true underlying image. However without a prior model, as discussed in Section 1.1, results are often artifact and noise heavy. In essence, the recovered image will vary drastically for each different realization of noise.

This dilemma is often referred to as the bias-variance trade-off [14]. Roughly speaking, variance can be thought of as the amount of noise and artifacts introduced into a reconstruction, and bias can viewed as systematic changes away from the true underlying image. In Chapter 3 we discuss how we try to alleviate one form of bias using data-driven priors. Additionally, in Chapter 5, by comparing results from many different image reconstruction algorithms with different priors, we can begin to identify the bias introduced by each prior/method.

2

Radio Interferometry and the Event Horizon Telescope

RADIO astronomy is the study of celestial sources at radio frequencies [124]. Imaging in radio frequencies allows us to see parts of our universe that we are not able to see naturally. For instance, Figure 2.1a shows an image of the galaxy Centaurus A in visible wavelengths seen by looking through an optical telescope. In the center is an image of the same location, but this time in radio wavelengths, imaged using a radio telescope [76]. By compositing the two images together we can see that the visible and optical wavelengths tell us very different stories about the galaxy's environment. Although we can see stars and dust in the optical image, the radio image shows us something that was once invisible, but very exciting – a jet of radio emission caused by a supermassive black hole at the center of the galaxy [113].



(a) (b) (c)

Figure 2.1. Images from the nearest radio galaxy, Centaurus A. (a) Optical Image, (b) Radio Image, (c) Composite Image. Notice that the galactic-sized radio lobes that are being emitted from the supermassive black hole at the center of the galaxy are invisible to the optical image (a), and can only be seen when imaged in the radio (b). [76]

Astronomical radio imaging has played a critical role in many scientific discoveries. For instance, imaging the cosmic microwave background has helped us to learn about the big bang and our early universe [104]. Although we have already been able to learn a lot from these images, astronomers are always looking to push the boundaries of imaging to further science. Recently, astronomers have been developing methods to increase the resolving power of radio telescopes in a quest to verify theories of general relativity [35, 36].

One-hundred years ago, Albert Einstein published his theory of general relativity (GR) [39]. In the years since then, scientists have found lots of evidence in support of GR [3, 15, 38]. However, one prediction of this theory, black holes, still have not been directly observed. Although there are predictions for what a black hole might look like [23, 94, 118], we have yet to actually take a picture of a black hole's event horizon. The Event Horizon Telescope (EHT) is an international project whose goal is to take the very first image of the immediate environment around a black hole [36, 37]. Realizing this goal would not only substantiate the existence of a black holes, but also aid in studying general relativity in the strong field regime [22, 68].

Observing structure on the scale of a black hole's event horizon requires imaging with radio wavelengths (roughly 1.3 mm) at an unprecedented ≈ 20 μ-arcsecond resolution [35, 36]. Astronomers believe this would make it possible to resolve Earth's closest supermassive black hole, Sagittarius A* (Sgr A*), which lies at the heart of our own Milky Way galaxy [36, 50]. This 20 μ-arcsecond resolution is much smaller than has previously been used to image radio sources [81]. As with cameras, a single-dish telescope is diffraction limited, making it very difficult to achieve this level of resolution. However, simultaneously collecting data from an array of telescopes, called an interferometer, allows us to overcome the single-dish diffraction limit through computational imaging methods [124].

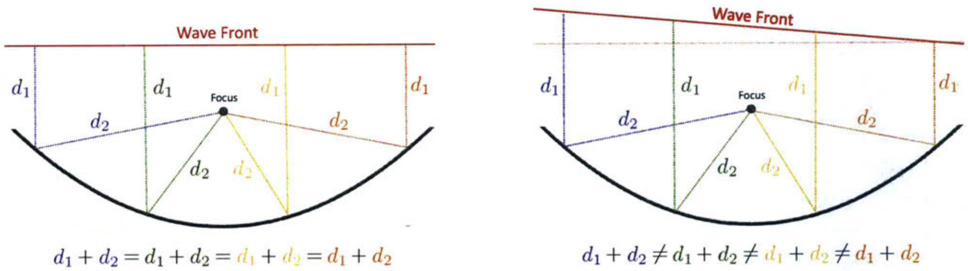
In this chapter we first discuss the limitations of single dish radio telescopes in Section 2.1. Next, we introduce, interferometry, a computational technique for high-resolution radio imaging in Section 2.2. These techniques are then extended to the case of disjoint telescopes distributed across the globe in Section 2.3. Following this we present a brief introduction of black holes and the current state of the Event Horizon Telescope in Section 2.4.

■ 2.1 Single Dish Telescope

The angular resolution of a single telescope depends on both the wavelength it operates at, λ, and on the diameter of the main dish (mirror), D [111]. The longer the wavelength, the worse the resolution, and the larger the diameter, the better the resolution. Consequently, a long-wavelength radio telescope has worse resolution than an optical- or infrared-wavelength telescope of the same size. In this section we briefly explain how this relation can be derived.

■ 2.1.1 Power Response

A traditional single-dish radio telescope acts like a giant reflector, or mirror, in order to gather light from the astronomical target. Light from the target source travels towards the telescope, appearing as a coherent wave parallel to the axis of the dish. This light is then reflected off the parabolic dish and is combined at the focal point. As the distance traveled from the incoming phase front to the focal point of the parabolic dish is the same for all rays, the reflected electric fields will all be in phase at the focus. The telescope's receiver response is a result of the coherent sum of the electric field at the focus [111]. Refer to the diagram in Figure 2.2a.



(a) Parallel Wave Front

(b) Angled Wave Front

Figure 2.2. A single-dish telescope uses a parabolic mirror to reflect and coherently collect light at a focal point. Due to the geometry of a parabola, light coming from a coherent wave front traveling parallel to the axis of the dish, will travel an equal distance to the focus from every reflected point (a). Thus, the electric field will be in phase, and constructively interference, at the focus. When the wave front is not traveling parallel to the dish, these distances will no longer be the same and the electric field will destructively interfere at the focus (b).

Alternatively, when the wave front comes in at an angle, the distance traveled from the wave front to the focal point is different for each ray. This causes the electric fields from this wave front to no longer be in phase at the focal point, resulting in destructive interference [111]. Refer to Figure 2.2b. In particular, when one side of the wave travels a distance $k \times \lambda$ more than the other side, then the sum of the electric field becomes 0. This occurs at angles approximately $k \times \frac{\lambda}{D}$ radians for $k = 1, 2, 3, \ldots$. When the phase differential across the full aperture is approximately $1.5\lambda, 2.5\lambda, 3.5\lambda, \ldots$ sidelobes appear in total received power. Each successive sidelobe is weaker than the last.

The power response of a single-dish telescope is given by

$$I = I_0 \left(\frac{2J_1(2\pi\theta D/\lambda)}{2\pi\theta D/\lambda} \right)^2, \tag{2.1}$$

where J_1 is the Bessel function of the first kind of order one, θ is the angular offset in radians, and I_0 is the maximum intensity [111]. A sample power response can be seen in Figure 2.3. This distinctive profile is often referred to as an Airy disk/pattern [111]. Notice that the first null occurs at $1.22\frac{\lambda}{D}$ radians. This figure also shows that most of the power is contained within the full width at half maximum (FWHM) of $1.02\frac{\lambda}{D}$ radians.

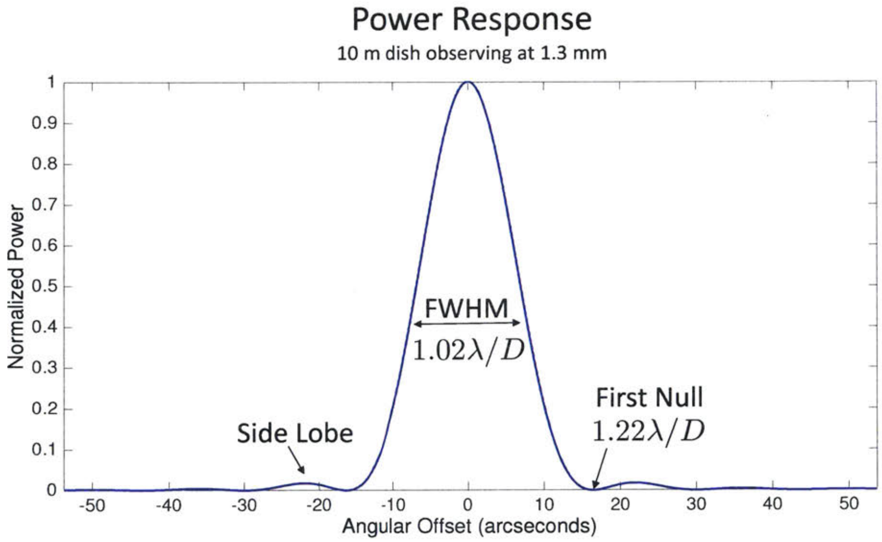


Figure 2.3. A sample power response profile for a 10 meter dish observing at a 1.3 millimeter wavelength. Note that most of the power is contained within λ/D radians of angular offset. This intrinsic property of parabolic reflectors leads to the fundamental relationship for single-dish resolution in Equation 2.2.

2.1.2 Angular Resolution

In Section 2.1.1 we discussed how most of the telescope's power is contained within roughly $\frac{\lambda}{D}$ radians. Therefore, it is not possible to resolve any structure smaller than this size. This is an intrinsic property of single-dish parabolic telescopes, and leads to the fundamental relationship

$$\theta_{rad} \approx \frac{\lambda}{D}, \tag{2.2}$$

where θ_{rad} is the smallest angle possible to resolve with a D diameter telescope observing at a λ wavelength (in common units) [111]. Through a change of units,

$$\frac{1 \text{ m}}{1e3 \text{ mm}} \times \frac{1 \text{ km}}{1e3 \text{ m}} \times \frac{180 \text{ deg}}{\pi \text{ rad}} \times \frac{60 \text{ arcmin}}{1 \text{ deg}} \times \frac{60 \text{ arcsec}}{1 \text{ arcmin}} \times \frac{1e6\ \mu\text{arcsec}}{1 \text{ arcsec}} \approx 2 \times 10^5,$$

we can rewrite this relation as

$$\theta_{\mu\text{arcsec}} \approx 2 \times 10^5 \times \frac{\lambda_{mm}}{D_{km}}. \tag{2.3}$$

2.2 Radio Interferometry

Imaging distant celestial sources with high resolving power (i.e. fine angular resolution) requires single-dish telescopes with prohibitively large diameters due to the inverse relationship between angular resolution and telescope diameter [124]. For example, it is predicted that emission surrounding the black hole at the center of the Milky Way subtends $\approx 2.5 \times 10^{-10}$ radians, or 50 μ-arcseconds [44]. Imaging this emission with a 20 μ-arcsecond resolution at a 1.3 mm wavelength would require a telescope with a roughly 13,000 km diameter.

Although a single telescope this large is unrealizable, by simultaneously collecting data from an array of telescopes, called an interferometer, it is possible to overcome the single-dish diffraction limit and create a virtual telescope as large as the maximum distance between telescopes in the array [124]. When these telescopes are distributed globally, this technique is referred to as *Very Long Baseline Interferometry* (VLBI) (see Section 2.3) [124].

An interferometer consists of P telescopes simultaneously observing and recording data-streams of light traveling from a common source: $\{V_n(t)\}_{n=1}^{P}$. By analyzing these data-streams, it is possible to obtain a sparse set of constraints on the spatial frequencies of the underlying source's emission image. These constraints can then be used to recover the underlying source emission image (see Chapter 3). In Section 2.2 we summarize the set of constraints obtained using interferometry, and in Section 2.2.1 we provide intuition behind these results.

The van Cittert-Zernike Theorem and Visibility Measurements

Formally, the *van Cittert-Zernike* theorem states that the time-averaged correlation of the measured signals from two telescopes, $\Gamma(u, v)$, is related to the ideal source image $I(\alpha, \delta)$ through a Fourier transform:

$$\Gamma(u, v) = R_c + iR_s \approx \int_\alpha \int_\delta I(\alpha, \delta) \exp\left[-i2\pi(u\alpha + v\delta)\right] d\alpha d\delta, \tag{2.4}$$

where (α, δ) is the angular sky coordinate in radians, and (u, v) is the dimensionless baseline vector between two telescopes, measured in wavelengths and orthogonal to the line of sight[1] [124]. Notice that Eq. 2.4 is just the Fourier transform of the source emission image, $I(\alpha, \delta)$. Thus, each pair of telescopes provides a single complex Fourier component of I at position (u, v) on the 2D spatial frequency plane. These measurements are referred to as *visibilities.* Since the spatial frequency, (u, v), is proportional to the baseline distance between telescopes, moving telescopes farther apart increases the resolving power of the interferometer, allowing it to distinguish finer details.

2.2.1 Intuition Under a Simple 1D Interferometer

We briefly describe a simplified view of interferometry in order to explain the relation described by the van Cittert-Zernike theorem. The goal of this section is to provide intuition; for additional details we recommend [124].

We consider a simple one-dimensional radio interferometer living in a two-dimensional world: two identical dishes separated by a vector distance B pointed at the same target source on the sky (e.g. a star, black hole, galaxy), in the unit-vector direction $\hat{s}$. Refer to Figure 2.4. Each telescope collects and records time-varying signals with frequency $\nu = \frac{c}{\lambda}$ (angular frequency $\omega = 2\pi\nu$). For ease of notation, we assume idealized electronics at each telescope with an isotropic antenna (e.g. uniform power response).

Astronomical emissions are generated by natural process. With the exception of rare cases (e.g. pulsars and masers), a telescope's received signal (with bandwidth $\Delta\nu$) is caused by the combination of short pulses of random occurrence, and has the form

---

[1] The change in elevation between telescopes can be neglected due to corrections made in preprocessing. Additionally, for a small field of view wide-field effects are negligible.

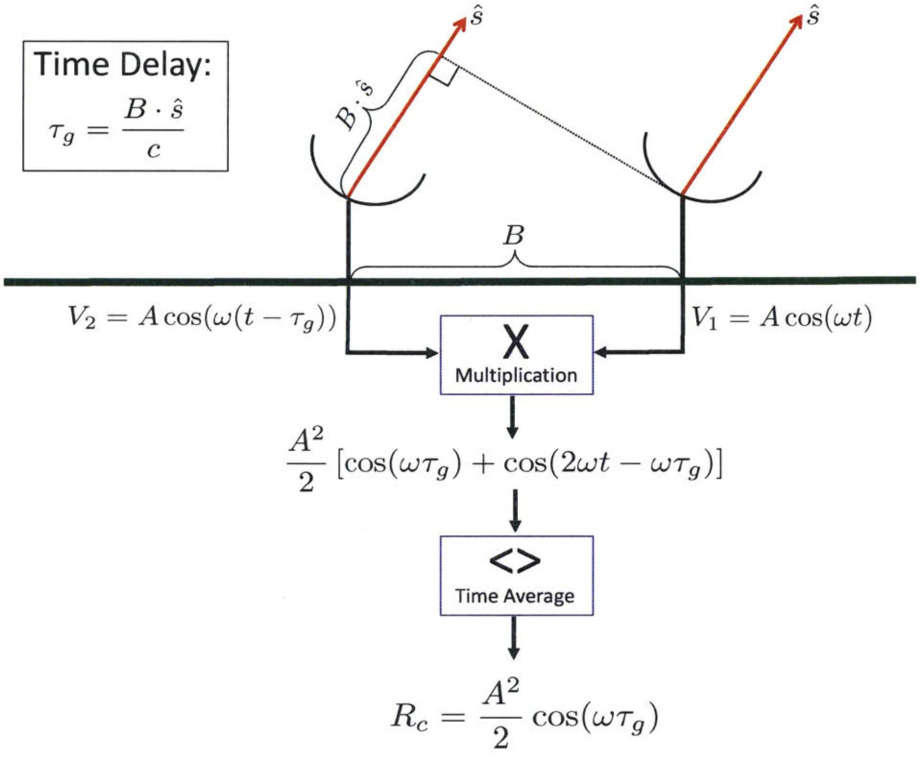


Figure 2.4. Simplified Interferometry Diagram: Light is emitted from a distant source and arrives at the telescopes as a plane wave in the direction $\hat{s}$. An additional distance of $B \cdot \hat{s}$ is necessary for the light to travel to the farther telescope, introducing a time delay between the received signals that varies depending on the source's location in the sky. The time-averaged correlation of these signals is a sinusoidal function related to the location of the source. This insight is generalized to extended emissions in the van Cittert-Zernike theorem and used to relate the time-averaged correlation to a Fourier component of the emission image in the direction $\hat{s}$.

of stationary Gaussian random noise [124]. However, for the purpose of simplification, we consider the unrealistic scenario of having a monochromatic source (i.e. a sinusoidal signal) that exhibits no propagation distortions due to the atmosphere.

Response of a Point Source

To begin, imagine that the target emission, in the direction $\hat{s}$, is simply a point source. In Section 2.2.1 we will extend the intuition gained from this simple scenario to the case of an extended emission on the sky.

Light travels from the far off source to the telescopes. However, because one telescope is slightly closer to the source than the other, there will be a time delay in the signals. As the source is very far away (i.e. the wave front is a plane wave), we can approximate the difference in the distance traveled to the telescopes as

$$\tau_g = \frac{B \cdot s}{c}, \tag{2.5}$$

where c is the speed of light. In other words, each received signal is a time shifted version of the other telescope's signal. From this *geometric time delay*, τ_g, the phase difference between the received signals can be calculated as

$$\phi = \omega\tau_g = 2\pi\frac{B \cdot \hat{s}}{\lambda}. \tag{2.6}$$

In particular, since we have assumed that we are observing a monochromatic source, the received signals at time t for telescopes 1 and 2 would be

$$V_1(t) = A\cos\left[\omega t\right] \tag{2.7}$$

$$V_2(t) = A\cos\left[\omega(t - \tau_g)\right] = A\cos\left[\omega t - \phi\right]. \tag{2.8}$$

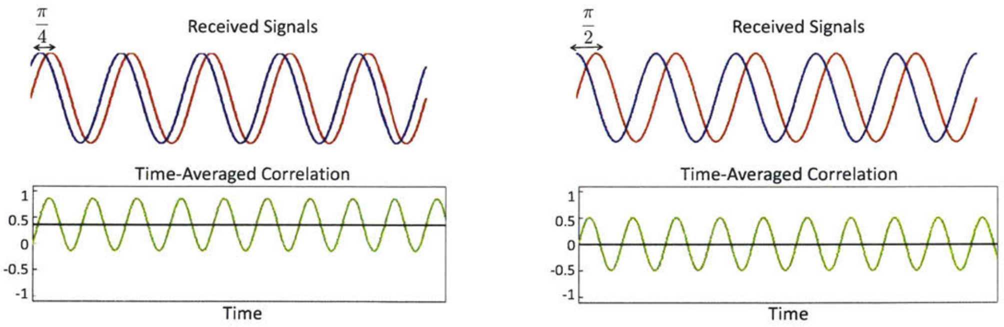
The time delay, τ_g, provides information about the precise location of a point source. Although we are unable to recover the precise value of τ_g, by taking the time-averaged correlation of V_1 and V_2, $\langle V_1(t)V_2(t)\rangle$, we are able to obtain a value that provides some information about the source's location. In particular, by using the trigonometric identity, $\cos(u)\cos(v) = \frac{1}{2}\left[\cos(u - v) + \cos(u + v)\right]$ when multiplying the time-varying signals $V_1(t)$ and $V_2(t)$, we find that

$$V_1(t)V_2(t) = \frac{A^2}{2}\left[\cos(\omega\tau_g) + \cos(2\omega t - \omega\tau_g)\right]. \tag{2.9}$$

Note that $V_1(t)V_2(t)$ is the sum of 2 terms: one that is unchanging and depends on τ_g, and one that is rapidly varying with time. By averaging over time the rapidly varying component is removed and we obtain an expression that is purely a function of observing frequency, ω, and the geometric time delay τ_g:

$$R_c \triangleq \langle V_1(t)V_2(t)\rangle = \frac{A^2}{2}\cos(\omega\tau_g). \tag{2.10}$$

This processes is referred to as *correlation*. Figure 2.5 shows an example of performing these operations on signals that have different phase offsets.



(a) Phase delay of $\pi/4$

(b) Phase delay of $\pi/2$

Figure 2.5. The effect of taking the time-averaged correlation of two signals that are separated by a time delay. The signals received by the two telescopes are shown in red and blue, respectively. By multiplying these signals we obtain the green curve. And by averaging the values in the green curve we obtain the time-averaged correlation, shown as the straight black line. This resulting time averaged correlation is related to the phase shift between the signals, but it does not provide information about the absolute time delay.

By plugging in the expanded expression for τ_g we see that R_c reduces to a function of the baseline distance between two telescopes and the location of the source on the sky:

$$R_c = I\cos\left(2\pi\frac{B\cdot\hat{s}}{\lambda}\right) = I\cos\left(2\pi\frac{B}{\lambda}\cos(\alpha)\right) = I\cos\left(2\pi u\cos(\alpha)\right), \tag{2.11}$$

where α is the angle between the baseline B and $\hat{s}$, $I = A^2/2$, and $u = \frac{B}{\lambda}$ is simply the number of wavelengths between the two telescopes. Note R_c is not a function of the time of the observation (provided the source is not variable and the baseline B is unchanging), the absolute location of the telescopes (provided the emission is in the far-field), or the actual phase of the incoming signal.

By plotting R_c as a function of $\theta = \frac{\pi}{2} - \alpha$, we visualize how R_c changes with the location of the source. Figure 2.6 plots this for $u = B/\lambda = 10$ and $u = 50$. Near $\theta = 0$ degrees the resolution is approximately $\frac{1}{u} = \frac{\lambda}{B}$. Therefore, we further approximate R_c as

$$R_c \approx I\cos(2\pi u\alpha). \tag{2.12}$$

Although, projection effects near the horizon cause this approximation to no longer hold for small u, it generally is negligible for the u typically used in interferometry.

Note for instance, that two telescopes located 10 meters apart and observing at a 1 mm wavelength would correspond with a u of 10000.

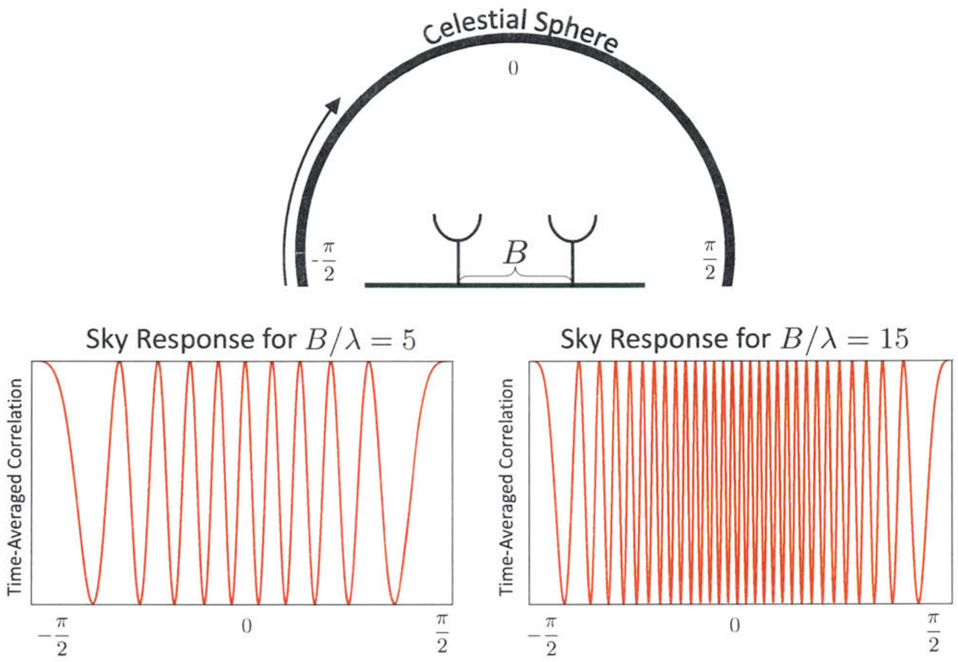


Figure 2.6. The time-averaged correlation for a single point source plotted as a function of its position on the celestial sphere. Notice that as the point source moves it sweeps out a sinusoidal pattern in the time-averaged correlation. The frequency of this sinusoidal pattern is a function of the distance between the two telescopes, B, and the observing wavelength, λ.

Response of an Extended Emission

The derivations shown in Section 2.2.1 all assumed that light was traveling from a point source relative to the resolution of the interferometer $\frac{1}{u}$. However, in the case of a spatially incoherent emission parameterized by the angle α, the resulting signals at the two telescopes become

$$V_1(t) = \int_{\alpha=-\pi/2}^{\pi/2} A(\alpha)\cos[\omega t + \theta(\alpha)]d\alpha \tag{2.13}$$

$$V_2(t) = \int_{\alpha=-\pi/2}^{\pi/2} A(\alpha)\cos[\omega(t + \tau_g(\alpha)) + \theta(\alpha)]d\alpha, \tag{2.14}$$

$$\tag{2.15}$$

where $\tau_g(\alpha) = \frac{Bcos(\alpha)}{c}$. Note that each telescope cannot recover the spatial distribution of A, as it is measuring a combined voltage. However, as in Section 2.2.1, combining information from both telescopes allows us learn more about the emission's structure. In taking the time-averaged correlation, $\langle V_1(t)V_2(t)\rangle$, we must average across all cross terms containing α_1 and α_2:

$$\begin{aligned}\langle V_1(t)V_2(t)\rangle &= \int_{\alpha_1}\int_{\alpha_2} \langle A(\alpha_1)\cos[\omega t + \theta(\alpha_1)]A(\alpha_2)\cos[\omega(t+\tau_g(\alpha_2)) + \theta(\alpha_2)]\rangle\, d\alpha_1 d\alpha_2 \\ &= \int_{\alpha_1}\int_{\alpha_2} \frac{A(\alpha_1)A(\alpha_2)}{2}\cos[\omega\tau_g(\alpha_2) - \theta(\alpha_1) + \theta(\alpha_2)]d\alpha_1 d\alpha_2. \qquad (2.16)\end{aligned}$$

Since the emission is spatially incoherent, the values of $\theta(\alpha_1)$ and $\theta(\alpha_2)$ are uniformly random and uncorrelated when $\alpha_1 \neq \alpha_2$. Therefore, this expression reduces to

$$R_c = \langle V_1(t)V_2(t)\rangle = \int_{\alpha} \frac{A(\alpha)^2}{2}\cos[\omega\tau_g(\alpha)] \approx \int_{\alpha} I(\alpha)\cos[2\pi u\alpha]. \qquad (2.17)$$

In other words, the time-averaged correlation of an extended emission is simply the projection of the emission's spatial power $I(\alpha)$ on a cosine, provided the emission is spatially incoherent. This cosine function is referred to as the *fringe pattern*, and its frequency, u, is determined by the baseline between the telescopes and the observing frequency through the relation $u = \frac{B}{\lambda}$.

The 1D van Cittert-Zernike Theorem

As explained in the beginning of Section 2.2, The van Cittert-Zernike theorem states that, by taking the time-averaged correlation, we can obtain a value, Γ, that is related to the ideal source image, I, through a Fourier transform. Mathematically, for a 1D interferometer,

$$\begin{aligned}\Gamma(u) &= \int_{\alpha} I(\alpha)\exp[-i2\pi u\alpha]d\alpha \qquad (2.18)\\ &= \int_{\alpha} I(\alpha)\cos(2\pi u\alpha)d\alpha - i\int_{\alpha} I(\alpha)\sin(2\pi u\alpha)d\alpha \qquad (2.19)\\ &= R_c - iR_s. \qquad (2.20)\end{aligned}$$

In Section 2.2.1 we showed how the first of these two terms, R_c, can be obtained through a time-averaged correlation. However, using only R_c would not make it possible to differentiate between the images I and $\frac{I(\alpha)+I(-\alpha)}{2}$. To correct for this, and determine the imaginary portion of Γ, another correlation is performed where a 90 degree phase shift is included in one of the signals – in effect, turning the cosine into a sine. Similar to Section 2.2.1, using the trigonometric identity, $\sin(u)\cos(v) = \frac{1}{2}[\sin(u-v) + \sin(u+v)]$ we obtain R_s:

$$R_s \triangleq \left\langle V_1(t) V_2\left(t + \frac{\pi}{2\omega}\right)\right\rangle = \int_\alpha \frac{A(\alpha)^2}{2} \sin[\omega \tau_g(\alpha)] = \int_\alpha I(\alpha) \sin[2\pi u \cos(\alpha)]. \tag{2.21}$$

2.2.2 Polarization

As explained above, correlating the received signals from a pair of telescopes provides information about the 2D spatial frequencies of the source image. However, typically the received signal at each telescope is broken up into two linear polarizations before it is recorded: left and right polarization. Depending on the signals used in the correlation, quantities related to either the total intensity or polarization of the light can be obtained. As the polarization direction often traces magnetic field lines in astronomical sources, reconstructing images of polarization allows us to characterize magnetic fields surrounding the source [52].

By correlating the common polarization direction from two telescopes (e.g. left-left or right-right) we obtain information about spatial frequencies in the total intensity image, often referred to as Stokes I. Alternatively, by correlating the opposite polarization direction (e.g left-right or right-left) we obtain information about the angle of polarization (Stokes Q and U parameters) [26, 70].

Since the image of linear polarization is a two-dimensional vector field, it can be represented as a complex image

$$P(\alpha, \delta) = Q(\alpha, \delta) + jU(\alpha, \delta), \tag{2.22}$$

where the real and imaginary parts of P are obtained by using the sum and the difference of the cross polarization terms, respectively. The primary difference between reconstructing polarization and total intensity is that the polarization Q and U images are not restricted to be positive [26]. Furthermore, the correlations from left and right cross terms are often much weaker than when correlating the same polarization direction for total intensity visibilities [70].

■ 2.2.3 Earth Rotation Synthesis

Interferometric measurements place a sparse set of constraints on the spatial frequencies of the underlying source image. In particular, as we showed in the previous section, each pair of telescopes provides information about a single 2D spatial frequency. This frequency is related to the 2D baseline vector connecting the two telescope sites from the direction of the target source. Thus, at a single time, for an array with P telescopes, at most $P \times (P-1)/2$ spatial frequencies are measured (as some telescopes may be below the horizon) [124]. This would result in only 15 measurements for an array of 6 telescopes. However, as the Earth rotates at 7.27×10^{-5} radians per second, the projected vector connecting each pair of telescopes, $(u, v) = B \cdot \hat{s}$, changes. This results in sampling additional spatial frequencies along elliptical paths in the frequency domain [112, 124]. Refer to Figure 2.7. Combining the different measurements taken as the Earth rotates is referred to as *Earth Rotation Synthesis.* Earth rotation synthesis is essential for building up enough measurements to constrain image reconstruction. If a source is static, the interferometric measurements taken over time as the Earth rotates all correspond to the same underlying image.

■ 2.3 Very Long Baseline Interferometry (VLBI)

When telescopes in an interferometric array are distributed globally, radio interferometry is referred to as *Very Long Baseline Interferometry* (VLBI). Although in theory VLBI is identical to radio interferometry, in practice they are very different [124]. In this special case of interferometry, the telescopes cannot be physically connected due to the large distances. Thus, data is recorded at each telescope site along with a precise time stamp, often generated by a hydrogen maser clock. This light, frozen in recordings, is then shipped to a common location so that the visibilities can be extracted through correlation.

Correlating VLBI data to produce a set of visibilities is much trickier than in traditional short-baseline radio interferometry. For example, the relative motion of each telescope site with respect to the source is different, resulting in a distinct Doppler shift for each received signal, $\{V_p(t)\}_{p=1}^{P}$. This causes the measured phase to drift as a function of frequency, and must be accounted for in correlation [124]. Additionally, although the hydrogen maser clock is very accurate on short time scales, over the course of a night's observation it may drift by a few tenths of a microsecond. Furthermore, as explained in Section 2.3.1, differences in the quickly-evolving atmosphere above each

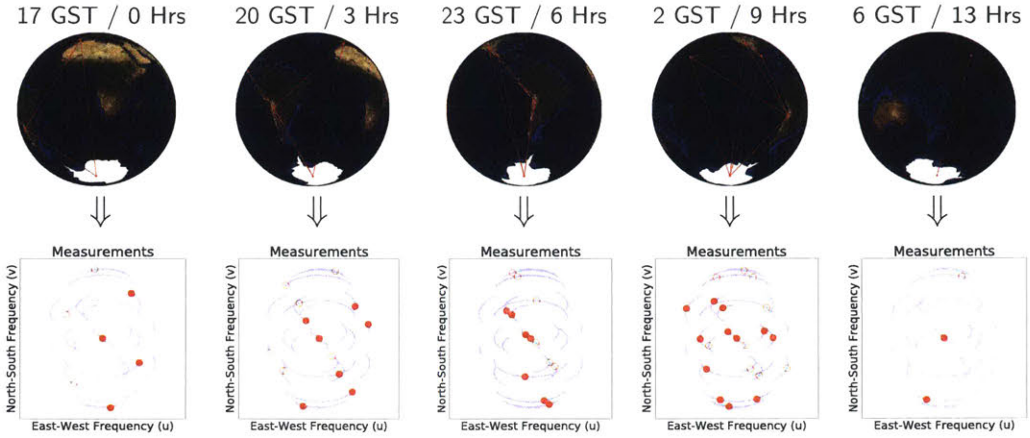


Figure 2.7. Earth Rotation Synthesis: For every two telescopes in the interferometric array, we obtain a single measurement (visibility) related to the underlying source image's 2D spatial frequency. This frequency is related to the baseline between the telescopes in the direction perpendicular to the observing source. This would be a prohibitively small number of measurements to make an image from. However, as the Earth rotates, the projected baselines change, and we observe new measurements related to different regions of the 2D frequency plane. As time increases (specified by the Greenwich Sidereal Time (GST)), the projected baselines change and the red dots on the lower frequency coverage plot, indicating the current measurements, change position. Assuming the source is static, this amounts to carving out elliptical paths in the frequency plane that are all related to the same image. This is visualized as the transparent blue lines. As light emitted from the source is real-valued we obtain two measurements on opposite sides of the frequency plane – each independent set of measurements displayed as either the open or closed red circles.

telescope cause propagation errors with a coherence time of only a few seconds. For these reasons, correlation must be done independently on small segments of $\{V_p(t)\}_{p=1}^{P}$ (typically 1 second segments for a 1.3 mm observation wavelength) before combining data to increase the signal-to-noise-ratio (SNR).

2.3.1 Instrument Noise & Calibration Errors

There are three sources of error on VLBI measurements that pose the most significant challenges from an imaging point of view: thermal noise, propagation error, and systematic errors due to amplitude miscalibration. The noise model assumed in this thesis for radio interferometry is given by Cornwell and Wilkinson [29] as:

$$\Gamma_{jk}^{\text{meas}} = \Gamma_{jk}^{\text{ideal}}(1 + a_j)(1 + a_k)\exp\left[i(\phi_j - \phi_k)\right] + \epsilon_{jk}. \tag{2.23}$$

For the true visibility $\Gamma_{jk}^{\text{ideal}}$ obtained using telescopes j and k, with amplitude error, a_j, phase error, ϕ_j, and thermal noise, ϵ_{jk}. In the following sections we describe each of these sources of error in detail.

Thermal Noise

Thermal agitation of the electrons in a circuit results in a small, yet detectable, current to flow, even when no voltage is applied. This introduces what is called thermal, or Johnson, noise into the measurements taken, and appears as isotropic Gaussian noise on each complex visibility, Γ [124]. Although the presence of this noise cannot be avoided, to help reduce its impact, receivers at each telescope are cryogenically cooled.

The standard deviation of thermal noise on a visibility is fixed based on bandwidth, $\Delta\nu$, integration time, τ, and each telescope's System Equivalent Flux Density (SEFD) for telescopes k and j.

$$\sigma_{jk} = \eta\sqrt{\frac{SEFD_k \times SEFD_j}{2 \times \Delta\nu \times \tau}} \tag{2.24}$$

The factor η results from losses due to quantization efficiency and is equivalent to 1/0.88 for 2-bit (4-level) quantization of the signal [124]. As the real and imaginary part of the visibility are correlated separately, the thermal noise should be zero-mean and isotropic. In other words, $\Re[\epsilon_{jk}], \Im[\epsilon_{jk}] \sim \mathcal{N}(0, \sigma_{jk}^2)$. Chapter 9 of [121] presents a derivation of Equation 2.24.

As noted above, the contribution of thermal noise from each telescope is expressed in by its SEFD. The SEFD specifies the flux density of a point source (in Jy) that would cause a doubling of the system's noise power [124]. Note that a lower SEFD value indicates higher sensitivity.

Atmospheric Propagation Error

All derivations in Section 2.2 assumed that light travels from the source to a telescope through a vacuum. However, in reality, differing atmospheres above each telescope site cause there to be large deviations in the relative propagation time [124]. In particular, for short radio wavelengths, fluctuations dominated by turbulence in the atmosphere's water vapor cause erratic changes in the path length light must travel to each telescope [92]. This path length affects an ideal visibility from telescopes k and j, $\Gamma_{k,j}$, by introducing an additional, rapidly varying, station-based phase term. Mathematically,

$$\Gamma_{k,j}^{\text{measured}} = \exp\left[i(\phi_k(t) - \phi_j(t))\right]\ \Gamma_{k,j}^{\text{ideal}}, \tag{2.25}$$

where $\phi_k(t)$ and $\phi_j(t)$ are time-varying phase delays introduced in the path to telescopes k and j, respectively. Equation 2.25 can be easily inferred by introducing these phase delays into $V_1(t)$ and $V_2(t)$ in Equation 2.8 and noting that $\langle V_1(t)V_2(t)\rangle$ then is equal to $\frac{A^2}{2}\cos(\omega\tau_g + \phi_2 - \phi_1)$ rather than $\frac{A^2}{2}\cos(\omega\tau_g)$.

For short radio wavelengths these phase offsets are stable for just seconds. In [92] the authors estimate the maximum coherent baseline, d_c[2] as $140 \times (0.001 \times \lambda)^{6/5}$ meters, when observing at a wavelength of λ meters. Under this approximation, the coherence time can be calculated for a frozen atmosphere with a constant wind speed, v_{wind}, as d_c/v_{wind} [92]. Table 2.1 shows a representative set of coherence times for different wavelengths when the wind speed is 10 meters per second under this model.

Wavelength	Coherence Time
10 cm	58 min
1.3 mm	19.2 sec
0.87 mm	11.8 sec
500 nm	4.4 ms

Table 2.1. An approximated coherence time for different wavelengths under a frozen atmosphere with a wind speed of 10 m/s.

Systematic Miscalibration Gain Error

Inaccurate estimation of each telescope's gain, as well as opacity fluctuations in the atmosphere, often cause there to be large miscalibrations in the reported visibility amplitudes. This is often due to the fact that calibration is much more difficult in VLBI than in traditional radio interferometry. Non-uniformity in the telescope dish/receivers, a lack of good point-like calibrators, varying atmospheres above each telescope site, and the fact that there are typically less telescopes in the array, all make the calibration process more difficult [124]. For this reason, there are often large systematic errors on the estimated visibilities amplitudes.

■ 2.3.2 Data Products

From the data-stream measurements taken at each telescope, $\{V_p(t)\}_{p=1}^{P}$, a number of different *data products* can be computed. This data is a function of the emission image's flux (brightness) distribution, and thus can be used to constrain image reconstruction.

---

[2] where the root mean square (RMS) phase fluctuations reach 1 radian

Depending on the imaging technique and the quality of the measurements, different data products may be used. In this section we review a number of common data products used in VLBI imaging, and throughout this thesis.

Complex Visibility

As explained in Section 2.2, each *visibility*, $\Gamma(u, v)$, is related to the ideal emission image through a Fourier transform. In particular, $\Gamma(u, v)$ extracts a complex 2D Fourier component of (u, v) cycles per radian from an image I with units of flux density [124]. As each visibility is calculated from a pair of telescopes, for an P telescope array we can obtain up to $\frac{P\times(P-1)}{2}$ visibility measurements at a single instant in time. Note that as the emission we observe is real-valued, we obtain two values on opposite sides of the frequency plane for each visibility measured: $\Gamma(u, v)$ and $\Gamma(-u, -v)$. For radio wavelengths, the noise appearing on perfectly calibrated visibilities is isotropic and caused by thermal effects (refer to Section 2.3.1).

Visibility Amplitude

Although atmospheric inhomogeneity causes substantial phase errors in a complex visibility, with careful calibration, the amplitude of the visibility can be mostly preserved. Constraining the amplitude of the visibilities in the case of atmospheric error can be very helpful in image reconstruction, especially when reconstructing an image with a large field of view. The *visibility amplitude* of visibility $\Gamma(u, v)$ is defined by:

$$|\Gamma(u,v)| = \sqrt{\Re\left[\Gamma(u,v)\right]^2 + \Im\left[\Gamma(u,v)\right]^2}. \tag{2.26}$$

As thermal noise is circularly Gaussian in the real-imaginary plane, a perfectly gain-calibrated visibility amplitude adopts the same standard deviation of noise as described in Section 2.3.1.

Bispectrum & Closure Phase

In Section 2.3.1 we explained how inhomogeneities in the atmosphere cause significant path length changes. These delays have a significant effect on the phase of measurements, and renders the phase unusable for image reconstructions at wavelengths less than 3 mm [92]. Although absolute phase measurements cannot be used, a clever observation allows us to still recover some information from the phases. The atmosphere affects an ideal visibility (spatial frequency measurement) by introducing an additional phase term: $\Gamma_{i,j}^{\text{meas}} = e^{i(\phi_i - \phi_j)}\Gamma_{i,j}^{\text{ideal}}$, where ϕ_i and ϕ_j are the phase delays introduced in the path to telescopes i and j respectively. By multiplying the visibilities from three different telescopes, we obtain an expression that is invariant to the atmosphere, as the unknown phase offsets cancel, see Eq. 2.29 [40].

$$\Gamma_{i,j}^{\text{meas}}\,\Gamma_{j,k}^{\text{meas}}\,\Gamma_{k,i}^{\text{meas}} = e^{i(\phi_i-\phi_j)}\Gamma_{i,j}^{\text{ideal}}\;e^{i(\phi_j-\phi_k)}\Gamma_{j,k}^{\text{ideal}}\;e^{i(\phi_k-\phi_i)}\Gamma_{k,i}^{\text{ideal}} \tag{2.27}$$

$$= e^{i(\phi_i-\phi_j+\phi_j-\phi_k+\phi_k-\phi_i)}\;\Gamma_{i,j}^{\text{ideal}}\,\Gamma_{j,k}^{\text{ideal}}\,\Gamma_{k,i}^{\text{ideal}} \tag{2.28}$$

$$= \Gamma_{i,j}^{\text{ideal}}\,\Gamma_{j,k}^{\text{ideal}}\,\Gamma_{k,i}^{\text{ideal}} \tag{2.29}$$

We refer to this triple product of visibilities in a closed loop as the *bispectrum* and the phase of the bispectrum as *closure phase.* The bispectrum is invariant to atmospheric noise; however, in exchange, it reduces the number of constraints that can be used in image reconstruction. Although the number of triple pairs in an P telescope array is $\binom{P}{3}$, the number of independent closure phases is only $\frac{(P-1)(P-2)}{2}$. For small telescope arrays, such as the EHT, this effect is large. For instance, in an eight telescope array, using closure phases rather than visibilities results in 25% fewer independent constraints on phase of the image [40]. Additionally, when using the bispectrum or closure phase we lose information related to the absolute location of the source.

Closure phases are often harder to interpret than complex visibilities. However, one very useful rule of thumb is that a closure phase of 0 or 180 degrees occurs if the source's emission image is symmetric [45]. To gain intuition behind this property, consider the Fourier transform of a double point-source separated by a distance (α_0, δ_0):

$$\mathcal{F}\left[\delta\left(\alpha - \frac{\alpha_0}{2}, \delta - \frac{\delta_0}{2}\right) + \delta\left(\alpha + \frac{\alpha_0}{2}, \delta + \frac{\delta_0}{2}\right)\right] \leftrightarrow cos\left[\pi(u\alpha_0 + v\delta_0)\right]. \tag{2.30}$$

Since the Fourier transform of a double point source is real, any visibility extracted from it will have a phase of 0 or 180 degrees (assuming no noise). As all symmetric images can be constructed from a sum of centered double point sources, any visibility from a symmetric image will also have a phase of 0 or 180 degrees. Therefore, the sum of three of these phases in a closure phase will always be 0 or 180 degrees, and anything other than this implies departure from symmetry.

Closure Amplitude

Similar to a closure phase, a closure amplitude is a data product computed from visibilities that is invariant to station-dependent amplitude errors [124]. Closure amplitudes can be computed using the visibilities found between a group of four telescopes:

$$\left|\frac{\Gamma^{\text{meas}}_{i,j}\Gamma^{\text{meas}}_{k,m}}{\Gamma^{\text{meas}}_{i,k}\Gamma^{\text{meas}}_{j,m}}\right| = \left|\frac{(1+a_i)(1+a_j)\Gamma^{\text{ideal}}_{i,j}(1+a_k)(1+a_m)\Gamma^{\text{ideal}}_{k,m}}{(1+a_i)(1+a_k)\Gamma^{\text{ideal}}_{i,k}(1+a_j)(1+a_m)\Gamma^{\text{ideal}}_{j,m}}\right| = \left|\frac{\Gamma^{\text{ideal}}_{i,j}\Gamma^{\text{ideal}}_{k,m}}{\Gamma^{\text{ideal}}_{i,k}\Gamma^{\text{ideal}}_{j,m}}\right|. \tag{2.31}$$

Although closure amplitudes are independent of all station-dependent amplitude errors they do not eliminate baseline-dependent errors that may be introduced in correlation [18].

2.4 The Event Horizon Telescope (EHT)

2.4.1 Testing General Relativity

In 1916 Albert Einstein published his famous theory of general relativity (GR) [39]. This theory described gravity as a curvature of a four-dimensional spacetime (3 space dimensions and 1 time dimension). In particular, it explains how massive objects, such as planets, stars, and black holes, deform spacetime and how this affects the path of other objects and light around them.

In the years since GR was first proposed, scientists have found a fair amount evidence in support of the theory. However, most of these experiments only probe the small spacetime curvature caused by smaller astrophysical objects, such as the Sun or the Earth[3] [15, 38]. This begs the question of whether this theory still holds in the extreme conditions around much more massive objects, such as a black hole, where the spacetime curvature is predicted to be extremely strong. Just as Newtonian physics was sufficient for relatively light objects, perhaps GR only holds around the objects that we have been capable of measuring thus far.

The Event Horizon Telescope (EHT) is an international project whose goal is to take the first image of the immediate environment around a black hole [35]. Realizing this goal would not only substantiate the existence of a black holes' event horizon, but also aid in studying GR in the strong field regime. Einstein's equations predict the size and shape of this ring (see Section 2.4.4). Thus, taking a picture would help to verify if GR holds in the extreme conditions around a black hole, where it is most likely to break down.

In order to take the first image of a black hole, astronomers predict we require a telescope with a resolution of roughly 20 μ-arcseconds for a wavelength of 1.3 millimeters (230 GHz). To achieve this with a traditional single-dish telescope would require an

---

[3] with the notable exception of LIGO [3]

Earth-sized dish. However, by using VLBI techniques, the EHT is building a virtual telescope capable of imaging at this extreme resolution [35].

■ 2.4.2 Targeted Black Holes

There are two primary black hole targets for the EHT [35]. These black holes are supermassive, meaning they reside at the center of galaxies and are millions to billions of solar masses in size [108]. Although these targets are not necessarily the closest of all known black holes to Earth[4], due to their massive size, their predicted event horizon appears the largest to us on the sky. As explained below, it is predicted that these black holes would appear as a small ring of light on the sky, roughly 20-50 μ-arcseconds in size. This is about the same size to us on Earth as an orange on the surface of the Moon.

Sagittarius A* (Sgr A*)

Near the border of the constellations Sagittarius and Scorpius is a bright compact astronomical source, with a total flux of roughly 2 Jy[5] [35]. This source, referred to as Sagittarius A* (Sgr A*), is predicted to be a supermassive black hole 26,000 light years from Earth at the heart of our own Milky Way galaxy [50].

Peering past galactic dust with infrared telescopes, astronomers have mapped the paths of a cluster of stars at the center of the Milk Way Galaxy for over 16 years. By tracking the motions of these stars over time, astronomers have concluded that the only thing small and dense enough to cause this motion is a supermassive black hole four million times the mass of the sun [50]. As explained in Section 2.4.4, this would result in a supermassive black hole with an event horizon of $GM/c^2 = 1.2 \times 10^{10}$ meters in size. By using the fact that a light year is 9.461×10^{15} meters in length, we can calculate the size of the event horizon on the sky as:

$$\tan^{-1}\left(\frac{1.2 \times 10^{10}}{26000 \times 9.461 \times 10^{15}}\right) = 4.9957 \times 10^{-11}\,\text{radians} \approx 10.3\,\mu\text{-arcseconds.} \quad (2.32)$$

As explained in Section 2.4.4 this would correspond to a ring of light of size 53.6 μ-arcseconds, detectable by the EHT's 20 μ-arcsecond resolution interferometer.

---

[4] stellar mass black holes, which are only a few solar masses in size, are scattered throughout our galaxy [64]

[5] right ascension (α): 17:45:40.041, declination (δ): -29:00:28.118

Messier 87 (M87)

At the head of Virgo resides the giant elliptical galaxy M87 – 54 million light years away[6]. Radio telescopes pointing towards the center of M87 have revealed a jet of emission caused by a supermassive black hole at the heart of the galaxy. If we could zoom in even further to the tip of this jet, astronomers predict that we would eventually see the event horizon of this supermassive black hole, glowing with a total flux of roughly 1 Jy [86]. Although the black hole in the center of M87 is much farther away from us, it is much larger than the black hole in the center of our own galaxy. Models predict that this black hole has a mass between 3 to 6 billion solar masses [47, 130], resulting in a event horizon of $GM/c^2 = 8.9 \times 10^{12}$ to 1.8×10^{13} meters in size. By using the same calculation as in Equation 2.32, we predict that this black hole would appear as a ring of light with diameter of 22.3 to 44.7 μ-arcseconds on the sky. Although M87's black hole is smaller and dimmer than Sgr A*, due to its massive size, it evolves on much longer timescales (see Section 2.4.4).

2.4.3 Observing Conditions & Equipment

In this section we briefly describe the equipment and operational conditions for the EHT.

Observing Wavelength

As explained in Section 2.2, one way to improve the EHT's resolution is by decreasing the observing wavelength. However, the light emitted from near the black hole's event horizon is primarily in the mm/sub-mm regime [48, 136]. Additionally, different wavelengths are affected differently by medium between the telescopes on Earth and the black holes they are observing. For instance, although visible wavelengths are shorter than radio wavelengths, they are not able to see into our galactic center, as their size causes them to interact with interstellar dust. However, just as humans cannot see through walls but radio waves are easily able to move through them, there exists larger wavelengths that can see through this galactic dust to the heart of our galaxy. In fact, there is only a small window of wavelengths around 0.87 to 1.3 mm that are able to pierce through to the core of the black hole and are not mostly absorbed by the atmosphere or heavily scattered by ionized interstellar medium [35]. Thus far the EHT has only observed at 1.3 mm (230 GHz). However, in the future the EHT plans to operate

---

[6] right ascension (α): 12:30:49.423, declination (δ): 12:23:28.044

at the shorter 0.87 mm wavelength (344 GHz) [42].

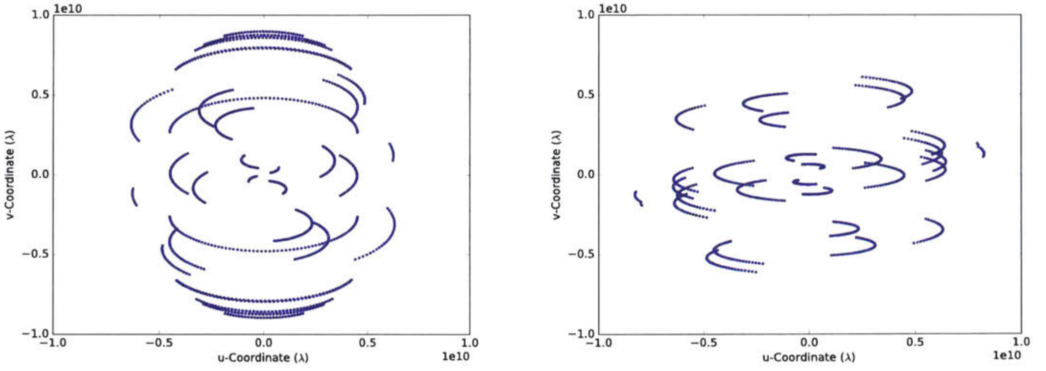
Telescopes and uv-Coverage

The EHT consists of an array of telescopes capable of observing in the millimeter to sub-millimeter wavelengths. In Table 2.2 we provide information about the array of the telescopes used in the EHT (in the present, past and near future). We provide the coordinates of each site along with its estimated system equivalent flux density (refer to Section 2.3.1). In the April 2017 campaign eight sites participated in the EHT array from six distinct locations: Hawaii, Arizona, Mexico, Chile, Spain, and the South Pole.

Station	Location	X	Y	Z	SEFD
CARMA*	California	-2397431.300	-4482018.900	3843524.500	3500
SMA	Hawaii	-5464588.447	-2492884.038	2150756.452	4000
JCMT	Hawaii	-5464584.676	-2493001.170	2150653.982	4700
SMT	Arizona	-1828796.200	-5054406.800	3427865.200	11000
LMT	Mexico	-768715.632	-5988507.072	2063354.852	1400
ALMA	Chile	2225061.873	-5440061.953	-2481682.084	100
APEX	Chile	2225039.530	-5441197.629	-2479303.360	3600
PV	Spain	5088967.900	-301681.600	3825015.800	1400
SPT	South Pole	792.600	-802.600	-6359569.200	9000
PDB**	France	4524000.430	468042.140	4460309.760	5200
GLT**	Greenland	1500692.0	-1191735.0	6066409.0	4744
KP**	Arizona	-1995678.840	-5037317.697	3357328.025	2500
HAY**	Massachusetts	1492420.4965	-4457272.10037	4296891.72893	2500

Table 2.2. **EHT Station Parameters:** The estimated SEFD (in Jy) and absolute X, Y, Z coordinates (in meters) relative to the center of the Earth associated with each station. * indicates a retired telescope. ** indicates a telescope that may be integrated into the EHT in the future.

As explained in Section 2.2, at a given time, each pair of telescopes produces a Fourier component of the source's emission image. The frequency of this measurement is related to the observing wavelength and the telescopes' baseline, in the direction perpendicular to the target source. By plotting out the location of these components as the Earth spins, we can visualize which frequency components are sampled by the EHT array for Sgr A* and M87 (refer to Figure 2.8). These plots are referred to as uv-coverage. The quality of the reconstructed images is highly dependent on the distribution of the uv-coverage.



(a) Sgr A*

(b) M87

Figure 2.8. The uv coverage of Sgr A* and M87 for the same telescope array observing over 24 hours. Telescopes in the 2017 EHT array observing at a 1.3 mm wavelength were used to generate these uv-coordinates. Notice that although the telescope array has remained unchanged for these two sources, the uv-coverage changes substantially. This is due to a difference in their source direction, $\hat{s}$.

Receiver and Backend Pipeline

Each telescope uses a heterodyne receiver to capture and record the light from the black hole onto hard drives. Light traveling from the black hole at 230 GHz (corresponding to 1.3 mm wavelength) is separated into its left and right polarization components before being mixed down to a manageable intermediate frequency (IF). The IF signal is then separated into two 2-GHz bands roughly corresponding to 228-226 GHz (lower side band) and 228-230 GHz (upper side band). Each polarization band's signal is then digitized using 2-bit quantization and recorded onto hard drives at a rate of 4096 MHz (in order to sample each 2-GHz band at the Nyquist rate). These recordings are then time stamped using an atomic hydrogen maser clock present at each telescope site. For a single polarization and band, one can calculate that the signal is being recorded onto hard drives at 1 gigabyte per second:

$$\frac{4096 \times 10^6 \text{ samples}}{1 \text{ second}} \times \frac{2 \text{ bits}}{1 \text{ sample}} \times \frac{1 \text{ gigabyte}}{8 \times 10^9 \text{ bits}} = 1 \; \frac{\text{gigabyte}}{\text{second}}. \qquad (2.33)$$

■ 2.4.4 Black Holes

A black hole is a region of spacetime that is so dense that its escape velocity exceeds the speed of light [114]. In this section we briefly review black holes to give an understanding of what the Event Horizon Telescope expects to see, and what properties effect a black hole visually.

The Schwarzschild Radius and No Hair Theorem

Black holes are thought to be formed by the collapse of stars greater than 2-3 solar masses ($M_{\odot}$), when a stars internal pressure can no longer resist gravitational pressure [25]. By equating kinetic energy to gravitational potential energy, one can calculate the radius necessary to compress a specified mass in order to create a black hole [114]. Mathematically,

$$\frac{v_{\mathrm{escape}}^2}{2} = \frac{GM}{R}, \tag{2.34}$$

for mass of M contained within a radius R, and Newton's gravitational constant $G = 6.6740810^{-11} m^3 kg^{-1} s^{-2}$. Solving for the radius where the escape velocity, v_{escape}, is equivalent to the speed of light, $c = 299,792,458$ m/s, results in the Schwarzschild radius, R_S, of the black hole, also known as its event horizon:

$$R_S = \frac{2GM}{c^2}. \tag{2.35}$$

In 1965 the Kerr-Newman metric was derived and showed that black holes are completely defined by just 3 parameters: their mass, spin and charge [98]. In other words, all other information about the matter that formed the black hole disappears behind the event horizon, and does not affect what an external observer sees. This theorem, coined the no-hair theorem (hair being a metaphor for details), fully describes how a black hole affects the nearby spacetime. In [8] it was shown that the apparent size and shape of the black hole depends mostly on the black hole mass, and to a much smaller degree, on its spin (the electric charge can be neglected for astrophysical black holes).

Accretion Disk, Inclination, and Doppler Beaming

Surrounding the black hole is a disk of spinning ionized gas (plasma) called an accretion disk [35]. This accretion disk is caused by gas that has accumulated close to the the black hole, but has yet to fall in. Contrary to popular belief, it is actually quite difficult for matter to fall into a black hole. If the gas is far enough away, gravity alone would not pull it in, and it would continue to orbit the black hole indefinitely, similar to planets around the Sun. However, friction resulting from particles in the disk rubbing up against each other leads to the gas to heating up. This causes the particles to lose energy and eventually fall into the black hole as well as brightly glow.

If we were to look at a black hole we would not be able to see the black hole itself, but instead a ring of light caused by the gravitational lensing of the bright accretion dish surrounding it. In other words, the black hole casts a "shadow" on a backdrop of bright material. The black hole's appearance changes depending on the viewing angle relative to the axis of the accretion disk. One reason for this is the fact that the immense gravitational field around the black hole causes light rays to bend differently depending on the viewing orientation [23]. However, even more striking, is the effect of a process called Doppler beaming (or alternatively, relativistic beaming). This process can cause the gas in the accretion disk to appear brighter or darker than it's intrinsic luminosity, $S_{\text{intrinsic}}$ because it is moving close to the speed of light [128][7]. If we were to view the accretion disk face-on, then the gas is not moving towards or away from us and we simply observe its intrinsic luminosity. However, if we view the accretion disk edge-on then some of the gas is moving towards us and some moves away from us. This causes a change in the observed luminosity of the gas, and results in an asymmetric emission image. In particular, if the emission source is moving away from Earth with a velocity v (if it is moving towards this value is negative), then for a resolved source

$$S_{\text{observed}} = S_{\text{intrinsic}} \left[\sqrt{\frac{1 - v/c}{1 + v/c}} \right]^{2-\alpha} \tag{2.36}$$

for spectral index α, where the spectral index is the slope on a diagram of $\log S$ vs. $\log \nu$ [78, 128]. In the frequency range the Event Horizon Telescope is targeting $\alpha \approx 0$. Thus, as the disk is moving towards us it appears much brighter than if it was static, and if it is moving away from us it is much fainter. Figure 2.9 shows what we predict a black hole may look like from different observation angles. Similarly, Figure 2.10 shows what we would predict an edge-on and face-on black hole would look like for varying spins.

Gravitational Lensing of the Last Photon Orbit

General relativity predicts that photons emitted by the gas spinning around a black hole travel along curved trajectories, forming a ring of light. Schwarzschild calculated the

---

[7] Here, we ignore the effect of redshift in $S_{\text{intrinsic}}$, that is due to the stretching of space-time as you get closer to the black hole's event horizion.


(a) 0° (b) 15° (c) 30° (d) 45° (e) 60° (f) 75° (g) 89°

Figure 2.9. The observed angle of inclination affects a black hole's appearance. Here we show a prediction of what a non-spinning black hole might look like at different angles ranging from face-on (0°) to edge-on (89°). [23]

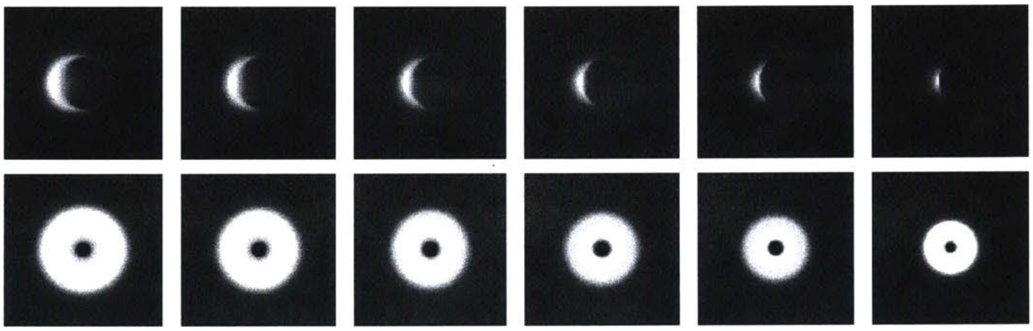

(a) 0.0 (b) 0.2 (c) 0.4 (d) 0.6 (e) 0.8 (f) 1.0

Figure 2.10. The spin of a black hole affects its appearance. Here we show a prediction of what a black hole might look like edge-on (top) and face-on (bottom) for a variety of different spins. 0.0 spin implies that the black hole is not spinning, while 1.0 spin implies it is maximally spinning. [23]

photon sphere at which photons are forced to travel in orbits for a non-rotating black hole as $\frac{3}{2}R_S$. However, due to lensing effects, to an observer on Earth, the apparent size of this sphere would be larger [36]. In fact, a sphere of radius R from a non-rotating black hole would appear at the size:

$$R_a = \begin{cases} \frac{3\sqrt{3}}{2}R & R \leq \frac{3}{2}R_S \\ \frac{R^{3/2}}{\sqrt{1-R_S}} & R > \frac{3}{2}R_S. \end{cases} \tag{2.37}$$

This indicates that the minimum apparent diameter of any light surrounding the black hole is $5.2R_S$ [36]. Thus, for a black hole with an event horizon of 10 μ-arcseconds, its event horizon and photon ring would appear as a larger ring of 52 μ-arcseconds to us on Earth.

Time-Variability

Although most large astronomical sources are static over the time scale of a night's observation, black holes' can have detectable structural changes on much shorter timescales [72, 87]. There are three primary physical phenomenon that cause this variability: a black hole's orbital velocity, magnetorotational instability, and interstellar scattering (when looking towards the Milky Way's galactic center). In Table 2.3 we provide a chart of the approximate timescales of each of these sources of variability for the EHT's targeted black holes, Sgr A* and M87.

Source of Variability	**Sgr A* :** $4 \times 10^6 M_\odot$	**M87:** $3 \times 10^9 M_\odot$	**M87:** $6 \times 10^9 M_\odot$
ISCO Orbital Period : 0% Spin	30.3 minutes	15.8 days	31.6 days
ISCO Orbital Period : 100% Spin	4.1 minutes	2.1 days	4.3 days
Magnetorotational Instability	$\approx$ minutes	$\approx$ days	$\approx$ days
Refractive Scattering	$\approx$1 day	N/A	N/A

Table 2.3. The level of variability induced by different physical phenomenon for Sgr A* and M87. The much larger mass of M87 makes it vary much more slowly than Sgr A*. [110]

Orbital Velocity For a rotating black hole with dimensionless spin value $a_{\text{spin}} \in [0, 1]$, the period of an orbit at radius r is given by

$$P_{\text{orbit}}(r) = 2\pi \left[\left(\frac{rc^2}{GM} \right)^{3/2} + a_{\text{spin}} \right] \frac{GM}{c^3}. \tag{2.38}$$

For a non-rotating black hole ($a_{\text{spin}} = 0$), this equation simply reduces to $P_{\text{orbit}}(r) = 2\pi\sqrt{r^3/GM}$ [72]. Note that for larger orbital radii, the orbital period increases.

The innermost stable circle orbit (ISCO) defines the smallest distance from a black hole where a particle with mass can orbit without falling in. The radius of this orbit depends on the spin of the black hole. For a non-rotating black hole ($a_{\text{spin}} = 0$) this radius is $6GM/c^2 = 3R_S$, and by substituting this radius into Equation 2.38 the orbital period of the ISCO is calculated as $2\sqrt{6^3}\pi GM/c^3$.

When a black hole is spinning (with maximum spin $a_{\text{spin}} = 1$ corresponding to the speed of light) the radius of the ISCO decreases. The black hole drags material along with it, a phenomenon referred to as frame dragging [114]. This causes stable orbits to come closer to the horizon. At a maximal spin of $a_{\text{spin}} = 1$ the ISCO radius is equivalent to the Schwarzschild radius and P_{orbit} reduces to $4\pi GM/c^3$.

Magnetorotational Instability (MRI) Magnetic fields within the accretion disk cause turbulence in a differentially rotating ionized gas (plasma) [7]. This comes about when magnetic fields try to tightly bind neighboring particles. However, as explained in the previous paragraph on orbital velocity, gas at different distances from the center of the black, r, hole will try to rotate around it at different speeds according to $\omega \propto 1/(r^{1.5}+a_{\text{spin}})$ [72]. In the case of a weak magnetic field, these two forces will fight against each other, resulting in the outer particles gaining angular momentum and moving further away from the black hole as nearer particles lose momentum and eventually fall in. This process causes the gas to destabilize and become turbulent.

Interstellar Scattering Ionized interstellar medium (ISM) in the line of sight between Earth and the galactic center causes scattering of the light coming from the black hole [69]. In particular, inhomogeneities in the interstellar medium cause refractive effects that impact the phase of light along the incident plane wave. This is very similar to viewing an image moving slowly behind frosted glass. The ISM is often modeled as a thin phase screen that moves slowly across the image over time. This process induces two types of scattering: diffractive and refractive scattering. Diffractive scattering causes a blurring effect that grows with the square of the observing wavelength. Refractive scattering causes image distortions and introduces substructure into the observed image [69].

■ 2.4.5 Preliminary Results

Although the EHT has yet to produce an image of a black hole, there are many significant results from observations made of Sgr A* using the instrument thus far.

First Detection of Structure In [36] Doeleman et al. reported to have detected structure in Sgr A* on similar scales as its lensed event horizon. As only three telescopes were used to make this observation, a picture could not be reconstructed. However, by analyzing the visibility amplitudes, it was determined that the emission had a angular diameter of roughly 37 μarcseconds. This smaller size relative to the predicted diameter (53 μarcseconds) may be due to observing the black hole edge-on relative to the accretion disk's axis. This would cause one side of the ring to appear much brighter than the other due to Doppler beaming, resulting in a smaller estimated angular diameter (refer to Section 2.4.4).

Detecting Asymmetry Asymmetry in the emission image was further predicted by Fish et al. in [45]. As discussed in Section 2.3.2, symmetric images should always produce closure phases with an angle of 0 or 180 degrees. Departures from this imply asymmetry. By analyzing closure phases from Sgr A* over 4 different years of observations, [45] concluded that there exists closure phases that show significant departure from 0 or 180 degrees. Not only does this imply structure exists in Sgr A* at the EHT's angular resolution, but also that this structure exhibits asymmetry. This could be explained by viewing the black hole edge-on, as proposed in [36].

Observing Variability In both [43] and [70] variability was detected in Sgr A*. In [43], Fish et al. analyzed 3 days of observation of Sgr A* and noticed that on the third observation night the total flux increased by roughly 17%. This new brighter flux remained stable for the rest of the observing night. Additionally, the measured intrinsic size of the source did not appear to change after the increase in total intensity. In [70] Johnson et al. reported having resolved partially ordered magnetic field structure and variability in the polarization direction near the event horizon of Sgr A*. In particular, by analyzing the variations in the polarization data products, and comparing them to the variation expected simply due to the Earths rotation, it was determined that Sgr A* must exhibit intrinsic variability in its magnetic fields.

3

Single Image VLBI Reconstruction of Static Sources

HIGH resolution celestial imaging is essential for progress in astronomy and physics. For example, achieving the Event Horizon Telescopes (EHT) goal of imaging the plasma surrounding a black hole's event horizon at high resolution could help answer many important questions; most notably, it may substantiate the existence of black holes [22] as well as verify and test the effects of general relativity [68]. However, imaging distant astronomical emissions, such as the black hole Sgr A*, with high resolving power would require a single-dish telescope with a prohibitively large diameter due to the inverse relationship between angular resolution and telescope diameter. Very long baseline interferometry (VLBI) alleviates the need for building an impossibly large single-dish telescope by simultaneously observing a common source from an array of telescopes distributed around the Earth. This technique makes it possible to emulate samples from a single-dish telescope with a diameter equal to the maximum distance between telescopes in the array, at the expense of having to handle missing data [124].

VLBI measurements place a sparse set of constraints on the spatial frequencies of the underlying source image (see Section 2.3.2 in Chapter 2) [124]. The task of reconstructing an image from these sparse constraints is highly ill-posed and relies heavily on assumptions made about the underlying image. Billions of dollars are spent on astronomical imaging systems to acquire the best measurements, yet traditional reconstruction techniques still suffer from unsophisticated priors and a lack of inverse modeling [109], resulting in sub-optimal images. Although traditional techniques have been reasonably successful in imaging large celestial sources with coarse angular resolution, they are quickly approaching their limits [85, 121].

The angular resolution necessary to see structure on the scale of a black hole's event horizon is believed to be at least an order of magnitude smaller than has been previously used to image radio sources [81]. Unfortunately, the difficulty of image reconstruction drastically increases as the angular resolution of a VLBI array improves. To improve angular resolution (i.e., increase resolving power), one must either increase the maximum distance between two telescopes or decrease the observing wavelength [124]. Due to the fixed size of Earth, increasing the maximum telescope baseline results in a smaller set of possible telescope sites to choose from. Therefore, algorithms must be designed to perform well with increasingly fewer measurements [85]. Extending VLBI to millimeter and sub-millimeter wavelengths to increase resolution requires overcoming many challenges, all of which make image reconstruction more difficult. For instance, at these short wavelengths, rapidly varying inhomogeneities in the atmosphere introduce additional measurement errors [92, 121].

Reconstructing an image using VLBI measurements is an ill-posed problem, and as such there are an infinite number of possible images that explain the data [85]. The challenge is to find an explanation that respects our prior assumptions about the "visual" universe while still satisfying the observed data. As measurements from the EHT become available, robust algorithms able to reconstruct images in the EHT's fine angular resolution regime will be necessary. In this chapter we present a new imaging method, CHIRP, developed specifically to handle the challenges faced in the EHT's data. We have found that these newly proposed methods can handle a wide range of different situations with minimal parameter tuning, and handle noise much better than other state-of-the-art imaging methods.

3.1 Previous Interferometric Imaging Approaches

VLBI image reconstruction has similarities with other spectral image reconstruction problems, such as Synthetic Aperture Radar (SAR), Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) [20, 89, 95, 123]. However, although measurements from VLBI share similarities with these other well-studied problems, VLBI image reconstruction faces a number of unique challenges. For instance, SAR, MRI, and CT are generally not plagued by large corruption of the signal's phase, as is the case due to atmospheric differences in mm/sub-mm VLBI[1]. Accounting for these differences is crucial in obtaining accurate astronomical images. Thus, imaging methods unique to VLBI must be developed. In this section we summarize a few significant algorithms from the astronomical interferometry imaging literature.

---

[1] In SAR the Fourier samples are all coherently related and the absolute phase can generally be recovered, even under atmospheric changes [65, 93]

■ 3.1.1 CLEAN

CLEAN is the de-facto standard method used for VLBI image reconstruction [60, 85]. This imaging method works under the assumption that the underlying source image can be described by using a discrete number of point sources. Below we discuss the procedure CLEAN uses to recover images [60, 121], some drawbacks of the method, and a general discussion. CLEAN is generally a very manual process, where at each iteration a knowledgeable user decides where to focus the method's attention [121]. However, in this thesis we do not discuss any tricks or techniques employed by CLEAN users, and focus solely on the core idea of the original method.

Procedure

CLEAN starts with an initial residual image, called the *dirty image*. The dirty image is computed by simply taking the inverse Fourier transform of the visibilities sampled by the interferometer. In other words, where a visibility is not sampled, the dirty image assumes a frequency value of zero. See Figure 3.1 for an example of a dirty image for the source image and uv-coverage shown.

From this initialization, CLEAN iteratively looks for the brightest point in the residual image and "deconvolves" around that location by removing side lobes that occur due to sparse sampling in the (u, v) frequency plane. In particular, at each iteration CLEAN selects the location of highest flux in the residual image and places a point source in a "final" image. It then subtracts the scaled point spread function (PSF) of the interferometer from the identified location in the residual image. The PSF, referred to as the *dirty beam*, is found by taking the inverse Fourier transform of the uv-coverage.

After many iterations, the residual image appears to just be noise. CLEAN then blurs the "final" image with a restoring beam to merge the point sources. The size of this anisotropic restoring beam is often determined by fitting a Gaussian to the *dirty beam* (in this case, the size is called the *nominal beam width*). Figure 3.1 shows the anisotropic nominal beam found by fitting a Gaussian to the dirty beam.

Self-Calibration

For mm/sub-mm wavelength VLBI, reconstruction is complicated by corruption of the visibility phases. CLEAN is not inherently capable of handling this problem since it uses complex visibilities. However, self-calibration methods have been developed to

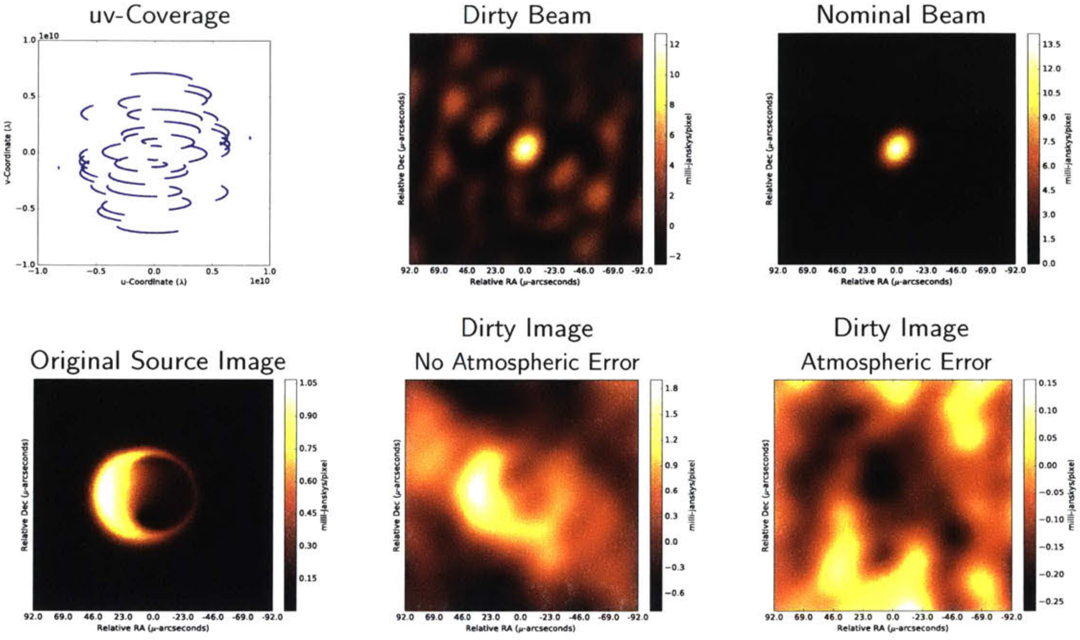


Figure 3.1. CLEAN Reconstruction Approach: The uv-coverage of an observation characterizes the PSF of the interferometer. The PSF, referred to as the dirty beam, is found by taking the inverse Fourier transform of the uv-coverage. The dirty image, the inverse Fourier transform of the complex visibility measurements, can be described as the original image convolved with the dirty beam. To recover an image, CLEAN iteratively removes sidelobes from the dirty image caused having a non-point like dirty beam. Afterwards, the final image is blurred by the nominal beam. This process only is appropriate when there is no atmospheric error in the visibility measurements. When atmospheric error exists, it is not possible to recover the underlying image simply by removing sidelobes. In this case a process called self-calibration is employed. The uv-coverage used in this example is generated according to the specifications of 3.4.1. Images shown are 64×64 pixels and the intensity of each pixel is specified in milli-janskys. The total flux of the original source image is 1 jansky.

greedily recover visibility phases during imaging [103]. These self-calibration methods work by iterating back and forth between CLEAN-ing the image and solving for closure constraints. See Section 2.3.2 in Chapter 2 for details on phase closure. Self-calibration requires manual input from a knowledgeable user and often fails when the SNR is too low or the source has a complex structure [121].

Discussion

CLEAN's approach to solving the inverse imaging problem is best described by the first approach discussed in Section 1.1 of the introduction (see Figure 1.1). CLEAN applies an inverse function (inverse Fourier transform), and then "cleans" up the result with post-processing. Although CLEAN implicitly contains a prior, by assuming the image is made up of point sources, this prior is not explicit and the trade-off between fitting the data and prior cannot be easily manipulated. In general the method is quite heuristic, which makes it very hard to adapt to situations when the data is not well-behaved, such as in millimeter-VLBI.

Although CLEAN is over 35 years old, theoretical understanding of the algorithm is limited. Success using CLEAN generally requires the use of many tricks and manual parameter tuning - especially in the case of self-calibration. Additionally, since it assumes a distribution of bright point sources, it struggles with reconstructing images containing extended emissions [121].

■ 3.1.2 Bayesian-Style Methods

Since CLEAN is not inherently capable of handing the atmosphere's corruption of the visibility phases in mm/sub-mm wavelengths, it frequently has trouble imaging with sparse, short-wavelength interferometric arrays like the EHT. However, more recently proposed Bayesian-style methods have made it possible to handle sparse and heterogeneous arrays in the mm/sub-mm regime, and can often even achieve some level of super-resolution in the images[2].

In order to reconstruct an image from the observed VLBI data, $\boldsymbol{y}$, these methods first approximate the continuous image $I(\alpha, \delta)$ as a square $M \times M$ array of values, $\boldsymbol{x}$, that represents the emission image's flux over a specified field of view (FOV). Using this representation, the imaging methods aim to solve

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}} \left[\chi(\boldsymbol{x}, \boldsymbol{y}) - \beta \mathcal{R}(\boldsymbol{x})\right] \tag{3.1}$$

where $\chi(\boldsymbol{x}, \boldsymbol{y})$ indicates how inconsistent the image, $\boldsymbol{x}$, is with the observed data, $\boldsymbol{y}$, and $\mathcal{R}(\boldsymbol{x})$ expresses how likely we are to have observed the image $\boldsymbol{x}$ [97]. As explained in Section 1.1.1 , these two terms often have different preferences for the "best" image, and fight against each other in selecting $\hat{\boldsymbol{x}}$. Their relative power in this decision is specified with the hyper-parameter β.

Multiple algorithms have taken this Bayesian-style approach to VLBI imaging.

---

[2] Super-resolution can be achieved by imposing priors that favor higher-frequency detail in the reconstruction than is captured by the interferometer. This is common when incorporating priors that favor sparsity, such as maximum entropy (MEM) priors [26]. However, MEM priors come at the cost of often reconstructing images that are sparser than the true image.

These algorithms often define a similar data inconsistency measure, $\chi(\boldsymbol{x}, \boldsymbol{y})$ (refer to Sections 2.3.2 and 3.2.2), but vary in what characterizes a "good" image, $\mathcal{R}(\boldsymbol{x})$. Maximum entropy and total variation priors have been used to construct $\mathcal{R}(\boldsymbol{x})$. These methods have been demonstrated in imaging optical interferometry data, and more recently in radio interferometry data taken with sparse telescope arrays [85].

Optical Interferometry

Interferometry at visible wavelengths faces the same phase-corruption challenges as mm/sub-mm VLBI. Although historically the optical and radio interferometry communities have been separate, fundamentally the resulting measurements and imaging process are very similar [92]. We have selected two optical interferometry reconstruction algorithms representative of the field to discuss and compare to in this work [85]. Both algorithms take the regularized maximum likelihood approach presented above, and can use the bispectrum, rather than visibilities, for reconstruction [9, 24].

BSMEM (BiSpectrum Maximum Entropy Method) takes a Bayesian approach to image reconstruction [24]. Gradient descent optimization [120] using a maximum entropy prior is used to find an optimal reconstruction of the image. Under a flat image prior BSMEM is often able to achieve impressive super-resolution results on simple celestial images. However, in Section 3.4 we demonstrate how it sometimes can over-resolve a source and introduce spurious detail.

SQUEEZE takes a Markov chain Monte Carlo (MCMC) approach to sample images from a posterior distribution [9]. To obtain a sample image, SQUEEZE moves a set of point sources around the field of view (FOV). The final image is then calculated as the average of a number of sample images. Contrary to gradient descent methods, SQUEEZE is not limited in its choice of regularizers or constraints [85]. However, this freedom comes at the cost of a large number of parameter choices that may be hard for an unknowledgeable user to select and tune.

■ 3.2 Defining the Forward Model & Conditional Distribution

Before discussing how to reconstruct images from VLBI data in Section 3.3, it is first important to explain how to evaluate the consistency of a proposed image, I, with the measured VLBI data, $\boldsymbol{y}$. To do this we present a recipe for constructing the forward model for a VLBI imaging system that relates a specified image to the data it would be expected to produce. As explained in Section 1.1 of the introduction, using this forward model one can then determine the conditional distribution of the data given an unknown image, to be used later in inference.

3.2.1 Deterministic Forward Model

Let $\boldsymbol{y}$ be a K-dimensional real vector of, possibly heterogeneous, data products. In order to evaluate how consistent a proposed image, I, is with the measured VLBI data, $\boldsymbol{y}$, we first define a function $f(I)$ to return a K-dimensional vector of ideal data products if I were the true underlying image. The deterministic forward model, $f(I)$, is composed of a set of sub-functions, $g_k(I)$, that each simulate an ideally observed data product:

$$f(I) = \begin{bmatrix} g_1(I) & g_2(I) & \ldots & g_K(I) \end{bmatrix}^T. \tag{3.2}$$

Each $g_k(I)$ produces a real-value corresponding to element k of the real-valued vector $\boldsymbol{y}$. For example, $g_k(I)$ may return the imaginary portion of I's spatial frequency, corresponding to a visibility obtained between a pair of telescopes. In the following subsections (3.2.1, 3.2.1, 3.2.1 and 3.2.1) we discuss how to approximate each sub-function $g_k(I)$ in VLBI data.

Continuous Image Representation

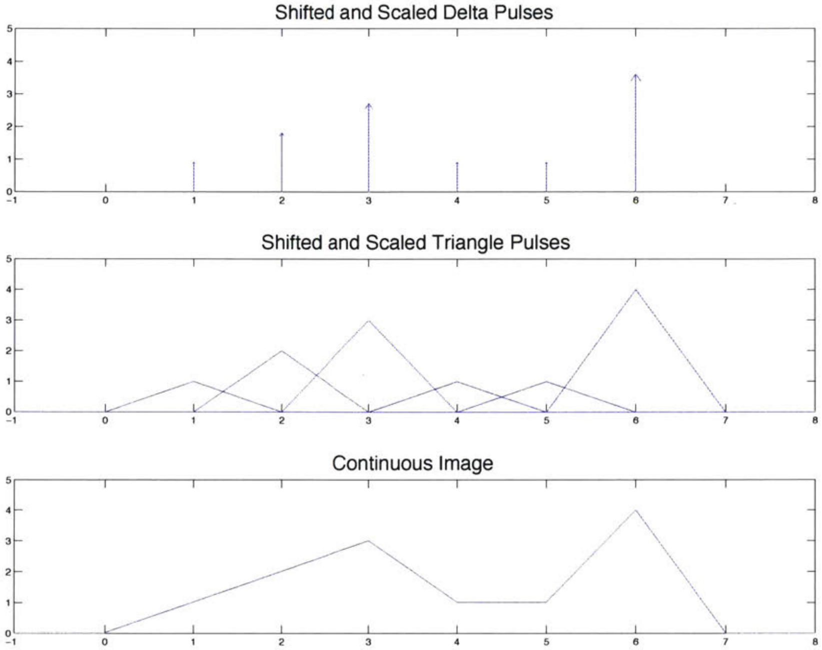
The image that we wish to recover, $I(\alpha, \delta)$, is defined over the continuous space of angular coordinates α and δ. Many imaging algorithms assume a discretized image of point sources during reconstruction [121]. This discretization either forces the reconstruction method to estimate a higher number of parameters than necessary, or introduces errors during the reconstruction process. Instead, we parameterize a continuous image using a discrete number of terms. This parameterization not only allows us to model our source emission with a continuous image, but it also reduces modeling errors during image reconstruction.

For a scene defined in the range $\alpha \in [-\frac{F_\alpha}{2}, \frac{F_\alpha}{2}]$ and $\delta \in [-\frac{F_\delta}{2}, \frac{F_\delta}{2}]$, we parameterize our space into $M \times M$ scaled pulse functions, $h(\alpha, \delta)$, centered around

$$\alpha = i\Delta_\alpha + \frac{\Delta_\alpha}{2} - \frac{F_\alpha}{2} \quad \text{for } i = 0, ..., M-1, \tag{3.3}$$

$$\delta = j\Delta_\delta + \frac{\Delta_\delta}{2} - \frac{F_\delta}{2} \quad \text{for } j = 0, ..., M-1, \tag{3.4}$$

for $\Delta_\alpha = \frac{F_\alpha}{M}$ and $\Delta_\delta = \frac{F_\delta}{M}$. Using Eq. 3.3 and 3.4 we then describe a continuous image


Shifted and Scaled Delta Pulses

Shifted and Scaled Triangle Pulses

Continuous Image

Figure 3.2. Image Representation: An example of a continuous 1D image defined in terms of 1D triangle pulses. Pulses are shifted and scaled by a discrete set of values, $\boldsymbol{x}$. These shifted and scaled pulses are then summed together to make a single continuous image. In this example $\boldsymbol{x} = [1, 2, 3, 1, 1, 4]$. Note how triangle pulses result in a piecewise linear continuous image.

as a discrete sum of shifted pulse functions scaled by $x[i, j]$. We refer to this image as $\hat{I}(\boldsymbol{x})$ for the M^2 vectorized coefficients $\boldsymbol{x}$. Mathematically,

$$\hat{I}(\boldsymbol{x}) = \sum_{i=0}^{M} \sum_{j=0}^{M} x[i, j] h \left(\alpha - \left(\Delta_\alpha i + \frac{\Delta_\alpha}{2} - \frac{F_\alpha}{2} \right), \delta - \left(\Delta_\delta j + \frac{\Delta_\delta}{2} - \frac{F_\delta}{2} \right) \right). \quad (3.5)$$

The chosen pulse places an implicit prior on the reconstruction. For instance, a sinc pulse with frequency and spacing Δ can reconstruct any signal with a bandwidth less than $\frac{\Delta}{2}$ [100]. In this work we choose to use a triangle pulse with width $(2\Delta_\ell, 2\Delta_m)$, since this is equivalent to linearly interpolating between pulse centers and also simplifies non-negativity constraints. Figure 3.2 shows how a discrete number of evenly spaced, overlapping triangle pulses can be used to linearly interpolate between the specified centers.

Extracting a Visibility

Since each measured complex visibility is approximated as the Fourier transform of $I(\alpha, \delta)$, the parameterization introduced in Section 3.2.1 above is especially convenient. Due to the shift theorem [100], substituting this image representation into the van Cittert-Zernike theorem (Equation 2.4) results in a closed-form solution to the visibilities in terms of $H(u, v)$, the Fourier transform of $h(\alpha, \delta)$. This process can be seen below:

$$\Gamma(u,v) \approx \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i2\pi(u\alpha+v\delta)}\hat{I}(\boldsymbol{x})d\alpha d\delta \tag{3.6}$$

$$= \sum_{i=0}^{N_\alpha-1}\sum_{j=0}^{N_\delta-1} x[i,j]e^{-i2\pi\left(u\left(\Delta_\alpha i+\frac{\Delta_\delta}{2}+a_\alpha\right)+v\left(\Delta_\delta j+\frac{\Delta_\delta}{2}+a_\delta\right)\right)}H(u,v) \tag{3.7}$$

$$= \mathbf{f}(u,v)^T\boldsymbol{x} = \left(\Re[\mathbf{f}(u,v)]^T + i\Im[\mathbf{f}(u,v)]^T\right)\boldsymbol{x}. \tag{3.8}$$

Note that performing a continuous integration has been reduced to a linear matrix operation with a $1 \times M^2$ complex row-vector $\mathbf{f}(u,v)^T$, similar to a Discrete Time Fourier Transform (DTFT).

If we choose to constrain our image reconstruction using visibilities, each $g_k(I)$ in $f(I)$ is defined as a linear function. In particular, as we have defined $\boldsymbol{y}$ to be composed soley of real-valued elements, if $\boldsymbol{y}_{2k} = \Re\left[\Gamma(u_k, v_k)\right]$ and $\boldsymbol{y}_{2k+1} = \Im\left[\Gamma(u_k, v_k)\right]$, then $g_{2k}(I)$ and $g_{2k+1}(I)$ can be approximated as

$$g_{2k}(\hat{I}(\boldsymbol{x})) = \Re[\mathbf{f}(u_k,v_k)]^T\boldsymbol{x} \tag{3.9}$$

$$g_{2k+1}(\hat{I}(\boldsymbol{x})) = \Im[\mathbf{f}(u_k,v_k)]^T\boldsymbol{x}. \tag{3.10}$$

In Figure 3.3 we show that this continuous image representation allows us to approximate the true frequency components more accurately than a discretized set of point sources, especially for high frequencies. Any pulse with a continuous closed-form Fourier transform can be used in this representation. This includes rectangle, triangle, sinc, and Gaussian pulses, as well as even a pulse related to cubic spline interpolation.

Diffractive Interstellar Scattering So far we have assumed that the visibility measurements simply provide a noisy measurement of the frequency component of the true image. However, in the case of observing Sgr A*, the frequency component we receive corresponds to an image that has been corrupted by interstellar scattering. As explained in Section 2.4.4, interstellar scattering consists of two components: diffractive

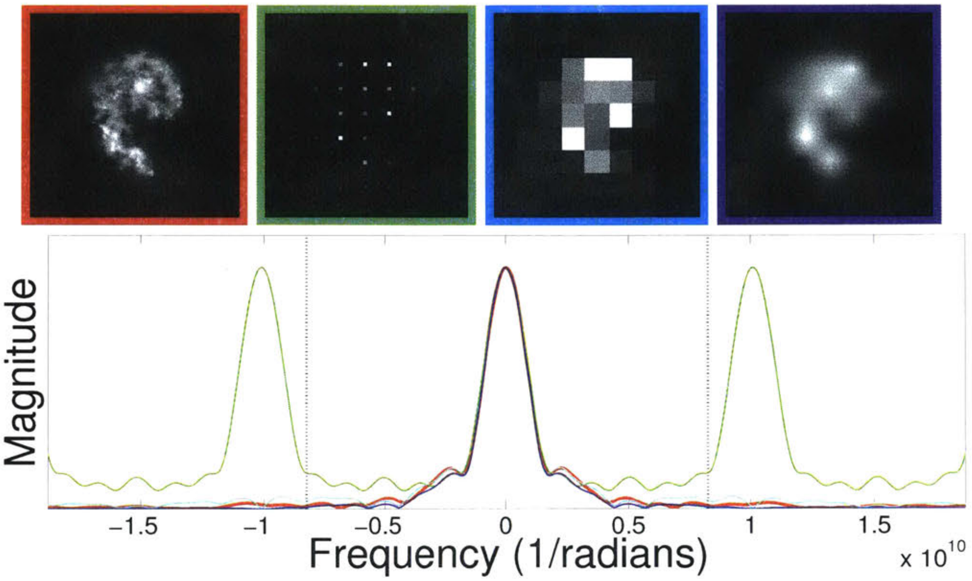


Figure 3.3. Modeling Error for Different Continuous Image Representations: Accurately modeling the frequencies of an image is crucial for fitting VLBI measurements during image reconstruction. Here we show, that with the same number of parameters, we can much more accurately model the true frequency distribution. A slice of frequencies for the true image is shown in red. Overlaid we show the effect of using the traditional discretized imaging model (green), and our improved model for rectangle (cyan) and triangle (blue) pulses. The dotted lines denote the frequency range sampled in the uv-coverage of Fig 3.1. Representing an image using triangle pulses reduces modeling errors for higher frequencies during image reconstruction.

and refractive scattering. Refractive scattering introduces substructure into the image that changes over time, similar to looking at an image under turbulent water. Diffractive scattering causes a blurring of the image, essentially convolving the true image with a Gaussian kernel [44, 69] Since refractive scattering changes over time, it is primarily described through its statistics and is not easily incorporated into the forward model. However, diffractive scattering can easily be incorporated in the forward model. See Appendix A.2 and our paper on scattering mitigation for more information on how to model and mitigate its effect [44].

Extracting Additional Data Products

Depending on the quality and type of data, we may wish to constrain the image reconstruction using different sets of data products. Although ideal visibility measurements obtained through VLBI correspond to 2D spatial frequencies, in the presence of corrupting atmospheric noise uniformly random phase errors are introduced into each complex measurement (see Section 2.3.1). In order to handle this additional phase error, without having to explicitly model the errors as latent variables, $\boldsymbol{y}$ may instead be populated with data products that are invariant to atmospheric inhomogeneity, such as the bispectrum, closure phase, visibility amplitude or closure amplitude data products.

A discussion of the data products used in image reconstruction can be seen in Section 2.3.2 of Chapter 2. All of these data products are a function of the visibilities. Therefore, each g_k associated with $\boldsymbol{y}[k]$ can be written in terms of the forward model for visibilities defined above. For example, each bispectrum measurement obtained from a closed loop of three telescopes, connected by three baselines corresponding to (u_1, v_1), (u_2, v_2), and $(-u_1 - u_2, -v_1, v_2)$, has the forward model:

$$B(\boldsymbol{x}, u_1, v_1, u_2, v_2) = \mathbf{f}(u_1, v_1)^T\boldsymbol{x} \times \mathbf{f}(u_2, v_2)^T\boldsymbol{x} \times \mathbf{f}(-u_1 - u_2, -v_1 - v_2)^T\boldsymbol{x}. \quad (3.11)$$

Therefore, if $\boldsymbol{y}[2k]$ and $\boldsymbol{y}[2k+1]$ are the real and imaginary components of a bispectrum measurement then, g is defined with the following polynomial equations:[3].

$$g_{2k}(\hat{I}(\boldsymbol{x})) = \Re\left[B(\boldsymbol{x}, u_1, v_1, u_2, v_2)\right] \quad (3.12)$$

$$g_{2k+1}(\hat{I}(\boldsymbol{x})) = \Im\left[B(\boldsymbol{x}, u_1, v_1, u_2, v_2)\right]. \quad (3.13)$$

■ 3.2.2 Conditional Distribution

The deterministic forward model $f(I)$ tells us what data we would expect to see in $\boldsymbol{y}$ under ideal circumstances when no noise is present in the system. However, a significant amount of noise is present on each data product (see Section 5.2.2). Additionally, as VLBI arrays are inhomogeneous, the level of noise on each measurement varies drastically. Therefore, we must incorporate how much we trust each data product when defining how inconsistent the image, $\hat{I}(\boldsymbol{x})$ is with the observed data, $\boldsymbol{y}$. To do this we define a function, $\chi(\boldsymbol{x}, \boldsymbol{y})$, related to the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$. In particular,

$$\chi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\sum_{k=1}^{K} \frac{(\boldsymbol{y}[k] - g_k(\boldsymbol{x}))^2}{\sigma[k]^2} \quad (3.14)$$

$$= \frac{1}{2}(\boldsymbol{y} - f(\boldsymbol{x}))^T\boldsymbol{R}^{-1}(\boldsymbol{y} - f(\boldsymbol{x})), \quad (3.15)$$

---

[3] Note that if interstellar scattering is included in each visibility this is incorporated into each $\mathbf{f}$

for $\boldsymbol{R} = \text{diag}[\sigma[1]^2, .., \sigma[K]^2]$ composed of the variance of noise on each $\boldsymbol{y}[k]$. Notice that in the case $p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R})$, $\chi(\boldsymbol{x}, \boldsymbol{y})$ is equivalent to $-\log p(\boldsymbol{y}|\boldsymbol{x}) + C$, for constant C.

Gaussian Noise Approximation for the Bispectrum

Although $p(\boldsymbol{y}|\boldsymbol{x})$ is Gaussian when $\boldsymbol{y}$ is composed fully of complex visibilities or visibility amplitudes (see Section 2.3.1), it is not Gaussian when $\boldsymbol{y}$ is composed of bispectra or closure phases, as each visibility is used to compute multiple terms. This means that $\chi(\boldsymbol{x}, \boldsymbol{y})$ is not truly the negative log-likelihood. However, in practice, we assume that each term of $\boldsymbol{y}$ is independent and can be described with a Gaussian noise model. In Appendix A.3 we discuss how we approximate the noise on each bispectrum. In particular, we approximate the variance of noise on each bispectrum $\Gamma_{i,j}^{\text{meas}}\Gamma_{j,k}^{\text{meas}}\Gamma_{k,i}^{\text{meas}}$ as

$$\sigma_{i,j}^2|\Gamma_{j,k}^{\text{meas}}\Gamma_{k,i}^{\text{meas}}|^2 + \sigma_{j,k}^2|\Gamma_{i,j}^{\text{meas}}\Gamma_{k,i}^{\text{meas}}|^2 + \sigma_{k,i}^2|\Gamma_{i,j}^{\text{meas}}\Gamma_{j,k}^{\text{meas}}|^2, \tag{3.16}$$

where $\sigma_{i,j}$ indicates the standard deviation of thermal noise on each visibility $\Gamma_{i,j}^{\text{meas}}$. By comparing our approximation to distributions obtained through sampling we have seen that it accurately models the true noise for SNR values greater than 1. As we expect values of SNR greater than 1 for imaging, a Gaussian noise model is a reasonable approximation for the bispectrum.

■ 3.3 Static Imaging using Data-Driven Priors

VLBI measurements are extremely sparse and noisy, particularly for the Event Horizon Telescope (EHT). Thus, since there are an infinite number of possible images that explain the data the reconstruction is ill-posed [85]. The challenge is to find an explanation that respects our prior assumptions about the "visual" universe while still satisfying the observed data.

In Section 3.2 we described how we define the consistency between a proposed image $\hat{I}(\boldsymbol{x})$ and the observed data, $\boldsymbol{y}$. To address the problem's ill-posedness, we must further narrow the space of possible solutions by defining a model that characterizes the appearance of a "good" image. However, it is not obvious what characterizes a "good" image, particularly in the case of the EHT. Since we have never seen a black hole before it is unclear what a likely black hole image should look like, and what we should assume about the structure of black holes. We could build an image prior using computer graphic renderings of black holes that assume General Relativity (GR) holds (such as are shown in Figures 2.9 and 2.10). However, doing this might cause some serious problems. For instance, what would happen if GR does not hold near a black hole? We would still want to reconstruct an accurate picture of the underlying image, but if we bake GR too much into our algorithms we may introduce significant bias in our reconstructions and just end up reconstructing what we expect to see.

An alternative to hand-designing regularizers is to design data-driven priors. This provides the flexibility to easily encode different image assumptions about our "visual" universe into the imaging process by simply training the image prior on different kinds of images. By imposing the features of different kinds of images, and seeing how the type of image we assume affects our reconstructions, we can begin to understand how much our imaging assumptions are biasing our final reconstructions. If all reconstructed images look roughly the same under different data-driven priors, then we can become more confident that the image assumptions we make are not significantly biasing our final picture of the black hole.

In Section 3.3.1 we describe an image reconstruction method developed under a trained multivariate Gaussian image prior. Although this simple image prior has not been previously explored in VLBI imaging, the intended goal of this section is primarily to gain intuition for the problem, not to introduce a novel imaging method. In Section 3.3.2 we extend these ideas to a new imaging method, called CHIRP, which uses a learned patch-based prior to reconstruct images under multiple "visual" universe assumptions.

■ 3.3.1 Gauss: Data-Driven Multivariate Gaussian Image Priors

In this section we describe an image reconstruction method developed under a trained multivariate Gaussian image prior. This simple formulation makes the problem simple to analyze and provides a clean interpretation of the results. In the results section we refer to this method as 'Gauss'.

Model

Representing I as its M^2 vectorized coefficients, $\boldsymbol{x}$, we define our observation model as:

$$\boldsymbol{y} \sim \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}) \tag{3.17}$$

$$\boldsymbol{x} \sim \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}), \tag{3.18}$$

where $\mathcal{N}_z(m, \Sigma)$ is the multivariate Gaussian distribution of z with mean m and covariance Σ. In this model, both the data likelihood, $p(\boldsymbol{y}|\boldsymbol{x})$, and the underlying image prior, $p(\boldsymbol{x})$, are multivariate normal distributions.

Training a Multivariate Gaussian Image Prior

We wish to learn the prior distribution from different sets of images. To do this, we collect a dataset of N images, parameterized by column-vectors $\{\boldsymbol{z}_n\}_{n=1}^{N}$. From this image set, the parameters of the multivariate Gaussian image prior can be found simply by taking the sample mean and covariance:

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{z}_n \tag{3.19}$$

$$\boldsymbol{\Lambda} = \frac{1}{N}\sum_{n=1}^{N} (\boldsymbol{z}_n - \boldsymbol{\mu})(\boldsymbol{z}_n - \boldsymbol{\mu})^T. \tag{3.20}$$

Optimization

Our goal is to find the most likely image, $\boldsymbol{x}$, under our model that describes the data products we have observed, $\boldsymbol{y}$. The posterior probability is written in terms the data likelihood, $p(\boldsymbol{y}|\boldsymbol{x})$, and the image prior, $p(\boldsymbol{x})$:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \tag{3.21}$$

$$= \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R})\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{3.22}$$

As explained in Section 1.1.1, a maximum a posteriori (MAP) solution is then found by minimizing the negative log-posterior:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} - \log p(\boldsymbol{x}|\boldsymbol{y}) \tag{3.23}$$

$$= \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[(f(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{R}^{-1} (f(\boldsymbol{x}) - \boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right]. \tag{3.24}$$

Note that this expression can be written in the common form

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[\chi(\boldsymbol{x}, \boldsymbol{y}) - \beta \mathcal{R}(\boldsymbol{x}) \right], \tag{3.25}$$

for $\mathcal{R}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$ and $\beta = -\frac{1}{2}$.

Linear Measurements As explained in Section 3.2.1, $f(\boldsymbol{x})$ is linear when $\boldsymbol{y}$ is composed solely of calibrated complex visibilities with no atmospheric error. In this case, $f(\boldsymbol{x}) = \boldsymbol{F}\boldsymbol{x}$, and a closed-form solution of $\hat{\boldsymbol{x}}$ can be found using the methods described in Section 1.1.1. In particular,

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}_{\boldsymbol{x}}(\hat{\boldsymbol{x}}, \boldsymbol{C}), \tag{3.26}$$

for .

$$\hat{\boldsymbol{x}} = (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F} + \boldsymbol{\Lambda}^{-1})^{-1} \left(\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{y} + \boldsymbol{\Lambda}^{-1} \mu\right) \tag{3.27}$$

$$= \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{F}^T (\boldsymbol{R} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1} (\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}), \tag{3.28}$$

$$\boldsymbol{C} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{F}^T (\boldsymbol{R} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1} \boldsymbol{F}\boldsymbol{\Lambda}. \tag{3.29}$$

In the limit of having no prior information about the underlying image $\boldsymbol{x}$, e.g $\boldsymbol{\Lambda} = \lim_{\lambda \to \infty} \lambda \mathbb{1}$, this MAP solution reduces to $\hat{\boldsymbol{x}} = \boldsymbol{F}^{-1}\boldsymbol{y}$. In other words, in the absence of prior image assumptions, the noise on each measurement, $\boldsymbol{R}$, is no longer relevant and the reconstructed image is simply obtained by inverting $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{y}$. When $\boldsymbol{y}$ are sparse visibilities, although F^{-1} is undefined, since $FF^{*T} = \mathbb{1}$, $\hat{\boldsymbol{x}} = \boldsymbol{F}^{-1}\boldsymbol{y}$ is very similar to reconstructing the dirty image.

Non-linear Measurements When $f(\boldsymbol{x})$ is a non-linear function of $\boldsymbol{x}$, as is the case when the data products in $\boldsymbol{y}$ are invariant to atmospheric noise, a closed-form solution does not exist. One way to try to solve for $\hat{\boldsymbol{x}}$ is through gradient descent. However, this method is slow. Alternatively, to solve for the optimal $\boldsymbol{x}$ we linearize $f(\boldsymbol{x})$ to obtain an approximate solution, $\hat{\boldsymbol{x}}$. Using a first order Taylor series expansion around $\tilde{\boldsymbol{x}}$, we approximate the data likelihood as

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}) \approx \mathcal{N}_{\boldsymbol{y}} \left(f(\tilde{\boldsymbol{x}}) + \dot{\boldsymbol{F}}(\boldsymbol{x} - \tilde{\boldsymbol{x}}), \boldsymbol{R} \right), \tag{3.30}$$

for $\dot{\boldsymbol{F}} = \left.\frac{f(\boldsymbol{x})}{d\boldsymbol{x}}\right|_{\tilde{\boldsymbol{x}}}$. Using this approximation, the optimal $\hat{\boldsymbol{x}}$ is

$$\hat{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\dot{\boldsymbol{F}}^T (\boldsymbol{R} + \dot{\boldsymbol{F}}\boldsymbol{\Lambda}\dot{\boldsymbol{F}}^T)^{-1} (\boldsymbol{y} + \dot{\boldsymbol{F}}\tilde{\boldsymbol{x}} - f(\tilde{\boldsymbol{x}}) - \dot{\boldsymbol{F}}\boldsymbol{\mu}). \tag{3.31}$$

A detailed derivation of $\dot{\boldsymbol{F}}$ for different data products is shown in Appendix A.4. To further improve the solution, we solve Equation 3.31 iteratively by updating $\tilde{\boldsymbol{x}}$ and setting $\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}}$ until convergence. Note that in the case that $f(\boldsymbol{x})$ is linear, $\dot{\boldsymbol{F}} = \boldsymbol{F}$ and Equation 3.31 reduces to Equation 3.28.

■ 3.3.2 CHIRP: Data-Driven GMM Patch Priors

Learning good image priors is a difficult task. Ideally we would like to learn the joint distribution of all pixels in an image. However, this is very challenging due to the exploding dimensionality and variability of images. Simple models, such as the Gaussian model presented in Section 3.3.1, can be trained. However, as we show in Section 3.4.2, these simple models are often not able to accurately characterize the manifold that the training images live on. This results in a poor image prior that incorrectly biases reconstructions. More complex models that more accurately characterize the distribution of images are prohibitively difficult to train or use in optimization[4].

Instead, a more manageable and successful, approach is to focus on modeling patches of the image, rather than the full image [138]. Although there may be a large number of possible configurations of the pixels in a patch, previous work has shown that many images contain patches that share very similar features [139]; for instance, there may be many patches that are of a constant intensity or contain a directional edge. In other words, the image patches can be well explained using a small set of building blocks. By reducing the dimensionality of the image data down to patches, we can use simpler models to accurately describe their distributions.

In this section we present a patch-based regularizer for VLBI imaging that models patches of an image using a Gaussian mixture model (GMM). We refer to our proposed method as CHIRP for Continuous High-resolution Image Reconstruction using Patch priors.

Model

As in Section 3.3.1, we approximate our observations model as

$$\boldsymbol{y} \sim \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}), \tag{3.32}$$

for the M^2 vectorized coefficients, $\boldsymbol{x}$, of $\hat{I}(\boldsymbol{x})$. However, unlike in Section 3.3.1, we define each overlapping $\widetilde{M} \times \widetilde{M}$ patch $\boldsymbol{z}$ of image $\boldsymbol{x}$ as being a sample from a Gaussian

---

[4] However, new neural network image generation models such as a variational auto encoders (VAEs) and generative adversarial networks (GANs) are beginning to show promise [66, 84]

mixture model (GMM) with C clusters. Mathematically,

$$\boldsymbol{z}_n = \boldsymbol{P}_n \boldsymbol{x} \sim \sum_{c=1}^{C} \boldsymbol{\pi}_c \mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c), \tag{3.33}$$

where matrix $\boldsymbol{P}_n$ extracts the n-th overlapping patch, $\boldsymbol{z}_n$, from $\boldsymbol{x} = x[i,j]$, and $\boldsymbol{\pi}_c$ is the mixture component weight of the c-th cluster, such that $\sum_{c=1}^{C} \boldsymbol{\pi}_c = 1$. From this relation we can write the probability of each patch $\boldsymbol{z}$ as

$$p(\boldsymbol{z}) = \sum_{c=1}^{C} \frac{\boldsymbol{\pi}_c}{\sqrt{|2\pi \boldsymbol{\Lambda}_c|}} \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_c)^T \boldsymbol{\Lambda}_c^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_c)\right). \tag{3.34}$$

Training a GMM Patch Prior

To train a Gaussian mixture model (GMM) patch prior we simply perform traditional EM optimization on a set of extracted patches [14]. In particular, given N patches $\{\boldsymbol{z}_n\}_{n=1}^{N}$, we iterate between the Expectation Step (E-step) of finding the expected likelihood of a patch belonging to each cluster,

$$\xi_c(\boldsymbol{z}_n) = \frac{\boldsymbol{\pi}_c \mathcal{N}_{\boldsymbol{z}_n}(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c)}{\sum_{k=1}^{C} \boldsymbol{\pi}_k \mathcal{N}_{\boldsymbol{z}_n}(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)}, \tag{3.35}$$

and the Maximization Step (M-step) of finding the model parameters that maximize the expected value of the log likelihood function:

$$\boldsymbol{\mu}_c = \frac{\sum_{n=1}^{N} \xi_c(\boldsymbol{z}_n) \boldsymbol{z}_n}{\sum_{n=1}^{N} \xi_c(\boldsymbol{z}_n)}, \quad \boldsymbol{\Lambda}_c = \frac{\sum_{n=1}^{N} \xi_c(\boldsymbol{z}_n)(\boldsymbol{z}_n - \boldsymbol{\mu}_c)(\boldsymbol{z}_n - \boldsymbol{\mu}_c)^T}{\sum_{n=1}^{N} \xi_c(\boldsymbol{z}_n)}, \quad \boldsymbol{\pi}_c = \frac{1}{N} \sum_{n=1}^{N} \xi_c(\boldsymbol{z}_n). \tag{3.36}$$

In this work we have set each $\{\boldsymbol{\mu}_c\}_{c=1}^{C}$ to 0 to force the GMM to learn symmetry in the relative intensities of patches in the images. Additionally, we use a patch size of 8×8 ($\widetilde{M} = 8$).

Optimization

We seek a maximum a posteriori (MAP) estimate of the parameters of the underlying emission image, $\boldsymbol{x}$, given the sparse observations, $\boldsymbol{y}$ and model parameters $\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c$, and $\boldsymbol{\pi}_c$. However, since we have defined the likelihood of an image patch, and not the likelihood of the image itself, we can no longer write a posterior distribution. Nevertheless, we still are able to write an optimizing cost function in the form

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[\chi(\boldsymbol{x}, \boldsymbol{y}) - \beta \mathcal{R}(\boldsymbol{x}) \right], \tag{3.37}$$

where

$$\mathcal{R}(\boldsymbol{x}) = \frac{1}{N} \sum_{n=1}^{N} \log \left[p(\boldsymbol{P}_n \boldsymbol{x}) \right] \tag{3.38}$$

Equation 3.37 appears similar to the familiar form using a Bayesian posterior probability; however, $\mathcal{R}(\boldsymbol{x})$ is not the log-likelihood of the full image, but instead we define it as the expected log likelihood of a patch in the image [138].[5]

To optimize Equation 3.37 we use "Half Quadratic Splitting" [138]. This method introduces a set of auxiliary patches $\{\boldsymbol{z}^n\}_{n=1}^N$, one for each overlapping patch $\boldsymbol{P}_n \boldsymbol{x}$ in the image, and a weighting parameter γ. We then solve this problem using an iterative framework:

(1) Solve for $\{\boldsymbol{z}^n\}_{n=1}^N$ given $\hat{\boldsymbol{x}}$: In order to complete this step we set $\{\boldsymbol{z}^n\}_{n=1}^N$ to the most likely patch under the prior, given the corrupted measurements $\boldsymbol{P}_n \boldsymbol{x}$ [138]. As this is hard to solve, we follow [138] and break this problem up into two manageable steps, that together approximate a MAP solution. Specifically, we first solve for the most likely cluster, $\hat{c}^n$, by evaluating the log-likelihood of each cluster:

$$\hat{c}^n = \underset{c}{\operatorname{argmax}} \left[\log \boldsymbol{\pi}_c - \frac{1}{2} \log |2\pi \boldsymbol{\Lambda}_c| - \frac{1}{2} (\boldsymbol{P}_n \hat{\boldsymbol{x}} - \boldsymbol{\mu}_c)^T \boldsymbol{\Lambda}_c^{-1} (\boldsymbol{P}_n \hat{\boldsymbol{x}} - \boldsymbol{\mu}_c) \right]. \tag{3.39}$$

Once we have selected the most-likely cluster we solve for the most likely patch under that cluster. We introduce a weighting parameter γ^{-1}, that indicates the variance of noise in the current estimate of $\hat{\boldsymbol{x}}$. Under the posterior distribution

$$p(\boldsymbol{z}^n | \boldsymbol{P}_n \hat{\boldsymbol{x}}; \boldsymbol{\mu}_{\hat{c}^n}, \boldsymbol{\Lambda}_{\hat{c}^n}) \propto \mathcal{N}_{\boldsymbol{z}^n}(\boldsymbol{P}_n \hat{\boldsymbol{x}}, \gamma^{-1} \mathbb{1}) \mathcal{N}_{\boldsymbol{z}^n}(\boldsymbol{\mu}_{\hat{c}^n}, \boldsymbol{\Lambda}_{\hat{c}^n}) \tag{3.40}$$

we estimate the best patch using Wiener filtering (as explained in Section 1.1.1):

$$\boldsymbol{z}^n = \boldsymbol{\mu}_{\hat{c}^n} + \boldsymbol{\Lambda}_{\hat{c}^n} (\gamma^{-1} \mathbb{1} + \boldsymbol{\Lambda}_{\hat{c}^n})^{-1} (\boldsymbol{P}_n \hat{\boldsymbol{x}} - \boldsymbol{\mu}_{\hat{c}^n}). \tag{3.41}$$

---

[5] Additionally, as discussed in 3.2.2, $\chi(\boldsymbol{x}, \boldsymbol{y})$ is not the log-likelihood of the data when $\boldsymbol{y}$ is composed of bispectrum or closure phase data products.

(2) Solve for $\hat{x}$ given $\{z^n\}_{n=1}^N$: Once we have solved for $\{z^n\}_{n=1}^N$, we then re-define the regularizer term $\mathcal{R}(\boldsymbol{x})$ as

$$\mathcal{R}(\boldsymbol{x}) = -\frac{1}{N}\sum_{n=1}^{N}\left[\frac{\gamma}{2}(\boldsymbol{P}_n\boldsymbol{x}-\boldsymbol{z}^n)^T(\boldsymbol{P}_n\boldsymbol{x}-\boldsymbol{z}^n) - \log p(z^n)\right], \tag{3.42}$$

and solve the optimization problem defined in Equation 3.37. If we were able to work with visibilities our problem would be quadratic in $\boldsymbol{x}$, and we could solve then for $\boldsymbol{x}$ in closed-form:

$$\hat{x} = \left(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F} + \frac{\beta\gamma}{N}\sum_{n=1}^{N}\boldsymbol{P}_n^T\boldsymbol{P}_n\right)^{-1}\left(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{y} + \frac{\beta\gamma}{N}\sum_{n=1}^{N}\boldsymbol{P}_n^T\boldsymbol{z}^n\right). \tag{3.43}$$

However, in the case of using a non-linear $f(\boldsymbol{x})$ we can either solve for $\hat{\boldsymbol{x}}$ using gradient descent or though linearizion. In particular, just as was shown in Section 3.3.1, we linearize the data likelihood (see Equation 3.30) to obtain an approximate solution in closed-form:

$$\hat{x} = \left(\dot{\boldsymbol{F}}^T\boldsymbol{R}^{-1}\dot{\boldsymbol{F}} + \frac{\beta\gamma}{N}\sum_{n=1}^{N}\boldsymbol{P}_n^T\boldsymbol{P}_n\right)^{-1}\left(\dot{\boldsymbol{F}}^T\boldsymbol{R}^{-1}(\boldsymbol{y} + \dot{\boldsymbol{F}}\tilde{\boldsymbol{x}} - f(\tilde{\boldsymbol{x}})) + \frac{\beta\gamma}{N}\sum_{n=1}^{N}\boldsymbol{P}_n^T\boldsymbol{z}^n\right). \tag{3.44}$$

Refer to Appendix A.4 for derivations of $\dot{\boldsymbol{F}} = \left.\frac{df(\boldsymbol{x})}{d\boldsymbol{x}}\right|_{\tilde{\boldsymbol{x}}}$ in the case of different data products.

(3) Iterate: Note that as $\gamma \to \infty$ patches $\boldsymbol{P}_n\boldsymbol{x}$ are restricted to be equal to their auxiliary patch $\boldsymbol{z}^n$. Thus as we increase γ the optimization under regularizers shown in Equations 3.38 and 3.42 converge. As suggested in [138], we iterate between steps (1) and (2) for increasing γ values of $1, 4, 8, 16, 32, 64, 128, 256$, and 512.

Multi-scale Framework In the case that the forward model $f(\boldsymbol{x})$ is non-linear, convexifying the cost function using a Taylor series expansion is only approximate. Thus, we slowly build up $\hat{I}(\boldsymbol{x})$ using a multi-scale framework that helps us to avoid local minima in the final solution. We initialize the image $\hat{\boldsymbol{x}}$ with small random noise centered around the mean flux density (average image intensity). Using this initialization, we iterate between solving for a new image $\hat{\boldsymbol{x}}$ and increasing the image's size using the discretized formulation of the image presented in Section 3.2.1. This framework allows us to find the best low-resolution solution before optimizing the higher frequency detail in the image. In this chapter, we initialize CHIRP reconstructions using a set of 20 × 20 pulses and slowly increase to 64 × 64 pulses over 10 scales. The same sized patch regularizer is used for each scale.

3.4 Results and Discussion

Measurements from the EHT have yet to become available. Therefore, we demonstrate the success of our algorithm, CHIRP, on a sample of synthetic examples and real VLBI measurements collected by the VLBA-BU-BLAZAR Program [74].

3.4.1 Synthetic Measurements

For image results presented in this chapter synthetic data was generated using realistic parameters for the EHT array pointed towards the black hole in M87. The ALMA, SMT, LMT, SMA, PV, PDB, SPT, GLT, and CARMA telescopes were used to generate 1.3 millimeter data over 12 hours of observation[6]. Refer to Table 2.2 for information about the location and sensitivity of these telescopes. Corresponding (u, v) frequency coverage is shown in Figure 3.1.

Visibilities were generated using the MIT Array Performance Simulator (MAPS) software [86]. Visibilities from MAPS are not generated in exactly the same manner as our forward model presented in Section 3.2. For instance, to generate a more realistic measurement, MAPS integrates over a wedge in the frequency domain defined by the observation-bandwidth and specified time-integration.

Thermal and atmospheric error is introduced into each of the visibility measurements (except when specified). As explained in Section 2.3.1, the standard deviation of thermal noise introduced in each measured visibility is characterized by standard deviation

$$\sigma = \frac{1}{0.88}\sqrt{\frac{SEFD_k \times SEFD_j}{2 \times \Delta\nu \times \tau}}, \tag{3.45}$$

for bandwidth $\Delta\nu = 4096$ MHz, integration time $\tau = 12$ seconds, and each telescope's System Equivalent Flux Density (SEFD).

The geometry of an array imposes an intrinsic maximum resolution on the image you can reconstruct from its measurements. The minimum recoverable fringe spacing (corresponding to the maximum frequency) for this configuration is 24.72 μ-arcseconds.

---

[6] While CHIRP was being initially developed CARMA was still operational.

Figure 3.4 shows the effect of filtering out spatial frequencies higher than the minimum fringe spacing on images with a field of view of 184 μ-arcseconds. This is done by convolving the true source image with the nominal beam shown in Figure 3.1. These images set expectations on what is possible to reliably reconstruct from the VLBI data.

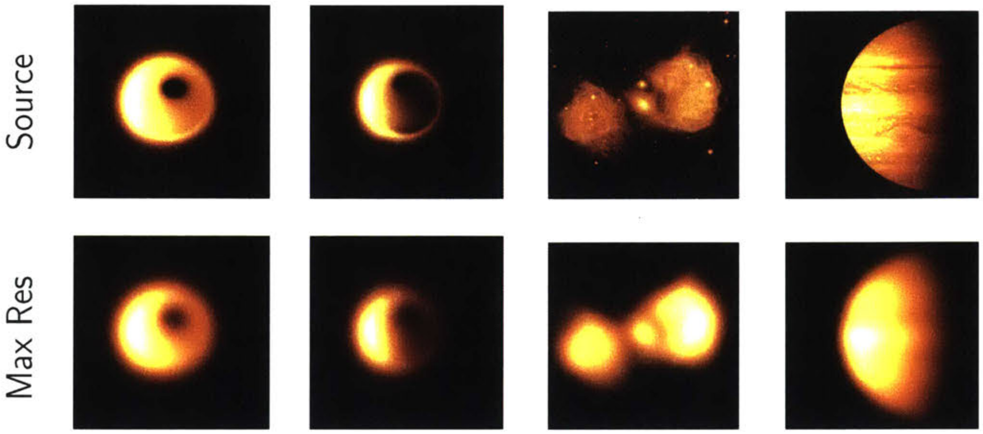


Figure 3.4. Intrinsic Maximum Resolution: The geometry of a telescope array imposes an intrinsic maximum resolution on images reconstructed from its measurements. Recovering spatial frequencies higher than this resolution is equivalent to super-resolution. For results presented in this chapter, the minimum recoverable fringe spacing (corresponding to the maximum frequency) is 24.72 μ-arcseconds. The original 'Source' images (183.82 μ-arcsecond FOV) are used to synthetically generate realistic VLBI measurements. We show the effect of filtering out spatial frequencies higher than the minimum fringe spacing for these source images in 'Max Res'.

■ 3.4.2 Bias Introduced Under Trained Image & Patch Priors

Flexibility of the data-driven prior framework allows us to easily incorporate a variety of different "visual" assumptions in our reconstructed image. For instance, in the case of the EHT, simulations of a black hole at different inclinations and spins can be used to train a model that can be subsequently used for reconstruction. We strive to develop a data-driven prior/regularizer that performs similarly across different trained models. This would indicate that the image assumptions being incorporated into the model are not overly image-type specific, and thus would not substantially bias reconstruction results in favor of a particular image structure. By training each model presented in Section 3.3 on different types of images, we can begin to understand how much bias each model imposes in the reconstruction process.

Training

A set of natural, astronomical, and black hole simulation images were collected and used to train the multivariate Gaussian image and GMM patch models. Refer to Figure 3.5. Training followed the procedures outlined in 3.3.1 and 3.3.2. When training the multivariate Gaussian image prior, images in each dataset were first resized to 64 × 64 pixels and assumed to have a field of view of 184 μ-arcseconds. Additionally, each of the natural and astronomical image datasets were augmented by rotating the images 90°, 180°, and 270°. For training the GMM patch prior, 8 × 8 pixel patches were randomly selected and extracted from each set of images. The number of clusters in each GMM patch model was selected empirically based on the size and variability of each dataset. Two-hundred clusters were used in the natural patch model, while thirty-two clusters were used in the astronomical and black hole patch models.

Note that the number of free parameters in the Gaussian image model is roughly $M^2 + M^4/2$. For $M = 64$ this is nearly 8.4 million free parameters, and far exceeds the amount of data that was used to train each model. Nonetheless, as we believe the true dimensionality of this data is much less than the number of free parameters, we are still able to gain insight into the characteristics of each trained model, and how each biases reconstruction results.

Visualizing the Models

To gain additional understanding into the structure that each set of images imposes through its prior, we display samples taken from each image and patch prior. Furthermore, we can visualize the principal components of each Gaussian model by computing the eigenvectors of $\mathbf{\Lambda}$ or $\mathbf{\Lambda}_c$. These eigen-emission images express the importance of different structures in each prior.

Image Model In Figure 3.6 we show four images obtained by drawing samples from the distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Lambda})$ for each image model. Additionally, below these samples we visualize the first 10 eigen-emission images for each model. Note how the model corresponding to each set of images is quite different. This can be seen by looking at both the image samples and the eigen-emission images. For instance, the natural image model produces bright flux regions that are uniformly distributed throughout the image. However, the astronomical image samples have larger flux confined primarily to the center of the image, resulting in eigen-emission images that have concentrated energy in the image center. This is due to the fact that most of the astronomical

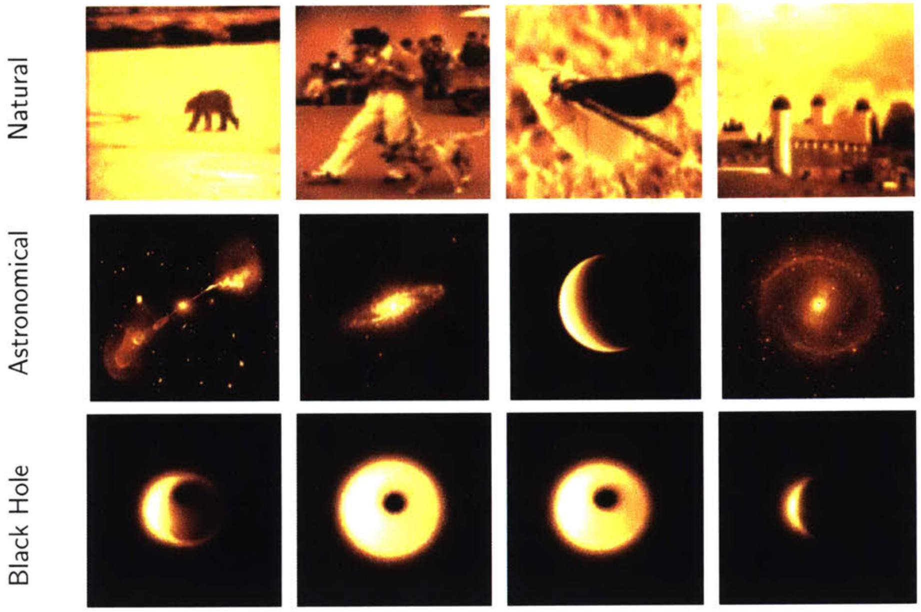


Figure 3.5. Training Image Datasets: The multivariate image prior and GMM patch prior were trained using using three distinct categories of images: natural (everyday), astronomical, and black hole simulation images. In this figure we show four random image samples from each dataset. Notice that each set of images has a very distinct look. A set of 400 natural images from the Berkeley segmentation dataset were used in the natural dataset [90]. Fifty (50) images from NASA/JPL-Caltech and NRAO were used in the astronomical dataset [2]. A set of 9000 computer generated images of Sgr A* for different inclinations and spins were used in the black hole dataset [23]. Additional images from this black hole dataset are shown in Figures 2.9 and 2.10.

images in the training set contain a bright center structure surrounded by the darkness of space. The black hole simulation model looks significantly different from the other two models. Since this model was trained using a very constrained set of images, the model focused on being able to capture small deviations in the location of the black hole shadow's sharp edge.

Patch Model In Figure 3.7 we show five patches sampled from the ten most likely clusters of each GMM patch model: $\{\mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c)\}_{c=1}^{10}$. Note that as we have set $\boldsymbol{\mu}_c$ to zero in each patch model, the captured structure is fully defined by $\boldsymbol{\Lambda}_c$, and inverting the intensity of a sampled patch is equally likely under the model. Similar features exist in all patch models. For instance, smooth directional edges appear in all three sets of sampled patches. Although these models share similarities, each also captures features special to its set of images. For instance, the model trained on astronomical images produces patches that contain dots, potentially indicating its attempt to capture the stars that appear in many of the astronomical images. Additionally, the black hole model generates many patches that contain curved edges that resemble the rim of light caused by the black hole's event horizon.

Comparing Model Bias

Figures 3.8 and 3.9 compare results obtained using the Gaussian image model and GMM patch model in Gauss and CHIRP, respectively. CHIRP results were obtained by constraining only the bispectrum data products (BIS) and using a multi-scale optimization approach as discussed in 3.3.2. Alternatively, we show results of Gauss when using complex visibilities (VIS), visibility amplitude and the bispectrum (AMP & BISP), and only the bispectrum (BISP). All datasets had the same thermal error included, but Gauss reconstructions done with complex visibilities did not contain atmospheric phase errors.

Image Model As image reconstruction is complicated by non-linear optimization in the case of atmospheric phase errors, the Gauss images reconstructed with complex visibilities best communicate the effect each trained prior has on the reconstructed image. Notice that, even in this much simpler situation, the images produced under each prior model differ substantially. The truth image in Figure 3.8 contains a picture of Jupiter that consumes the full field of view. Although this can be categorized as an astronomical image, the natural image prior produces the closest reconstruction to the true source image. This partly occurs because the natural image prior does not have a preference to only reconstruct flux in the center of the image, as is the case in the astronomical image prior. On the other hand, notice how in Figure 3.9 the astronomical image prior is able to pull out features of the true source image better than the natural image prior since most of the flux is concentrated in the center of the image. As expected, the black hole prior performs very poorly in Figure 3.8 since the brightness distribution in the source image is very different from the black hole images it saw in training. However, perhaps surprisingly, the black hole model also performs very poorly in Figure 3.9 despite the truth image being a sample from the same simulation model. This is due to the fact that the multivariate Gaussian model is not able to accurately model the manifold of black hole images, resulting in overfitting with negative flux during optimization. By analyzing these results it is clear that training a full image model on a specific set of images could substantially bias reconstructions.

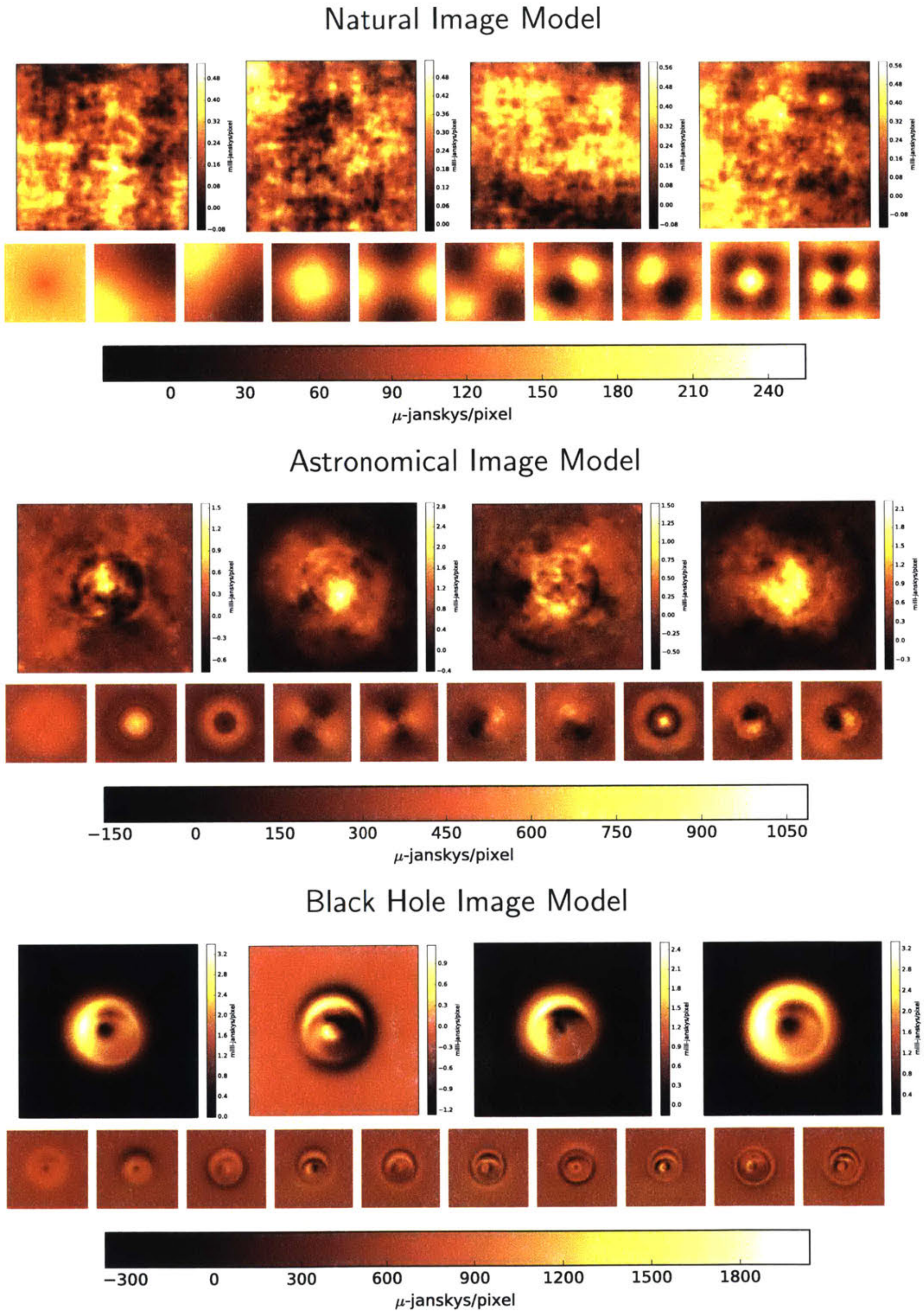


Figure 3.6. Trained Image Model: Four images are sampled from each of the trained Gaussian image priors. These are shown below each model name. The top ten eigen-emission images corresponding to each model are shown directly below the image samples. Each image model captures very different characteristics from its set of images. For instance, unlike in the natural image model, the astronomical image model concentrates most of its energy in the center of the image.

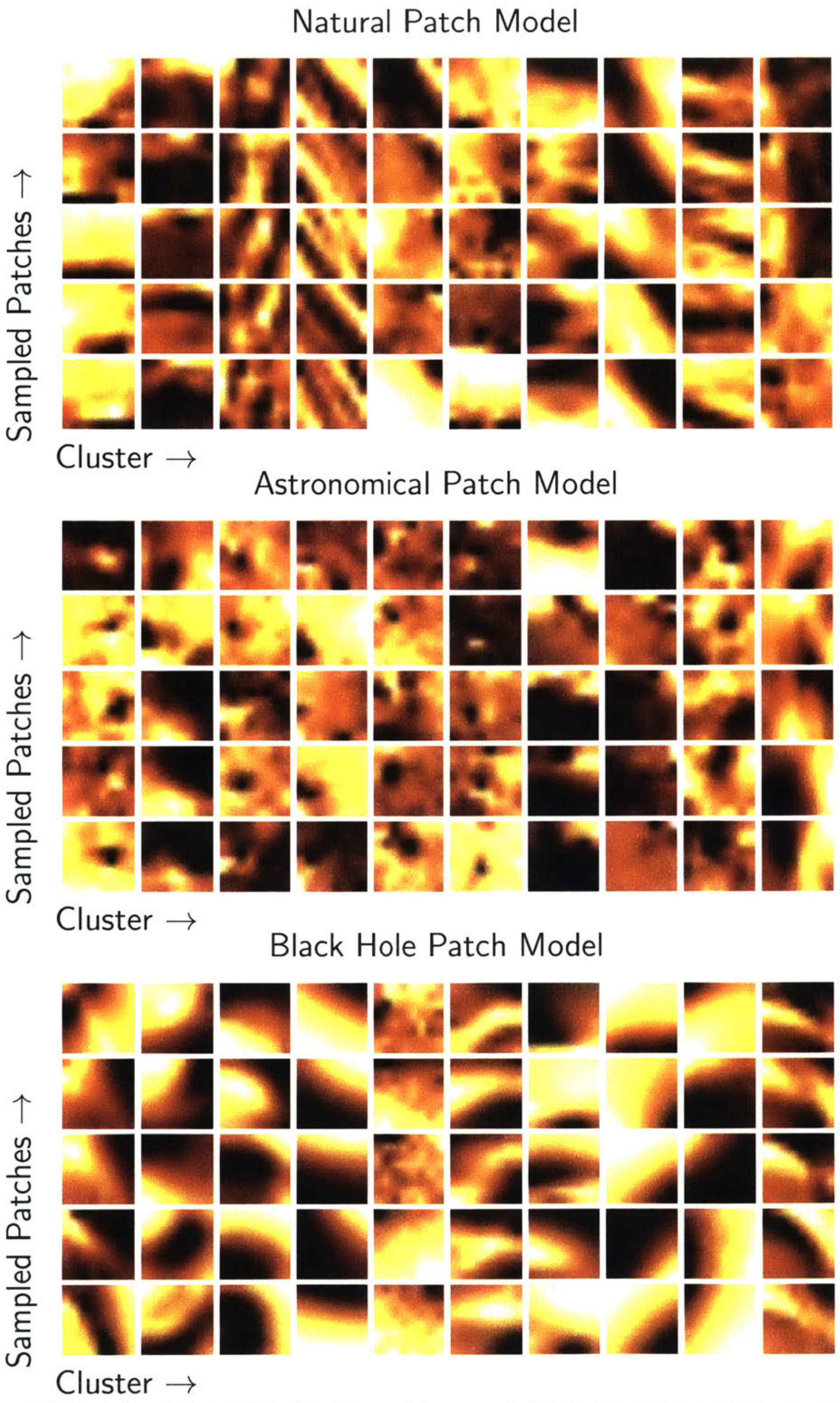


Figure 3.7. Trained Patch Model: Five patches sampled from the ten most likely clusters of each GMM patch model (left to right corresponds to more to less likely). Although similar features exist in all three patch models, there are also significant differences. Each patch model appears to capture characteristics specific to its own set of images.

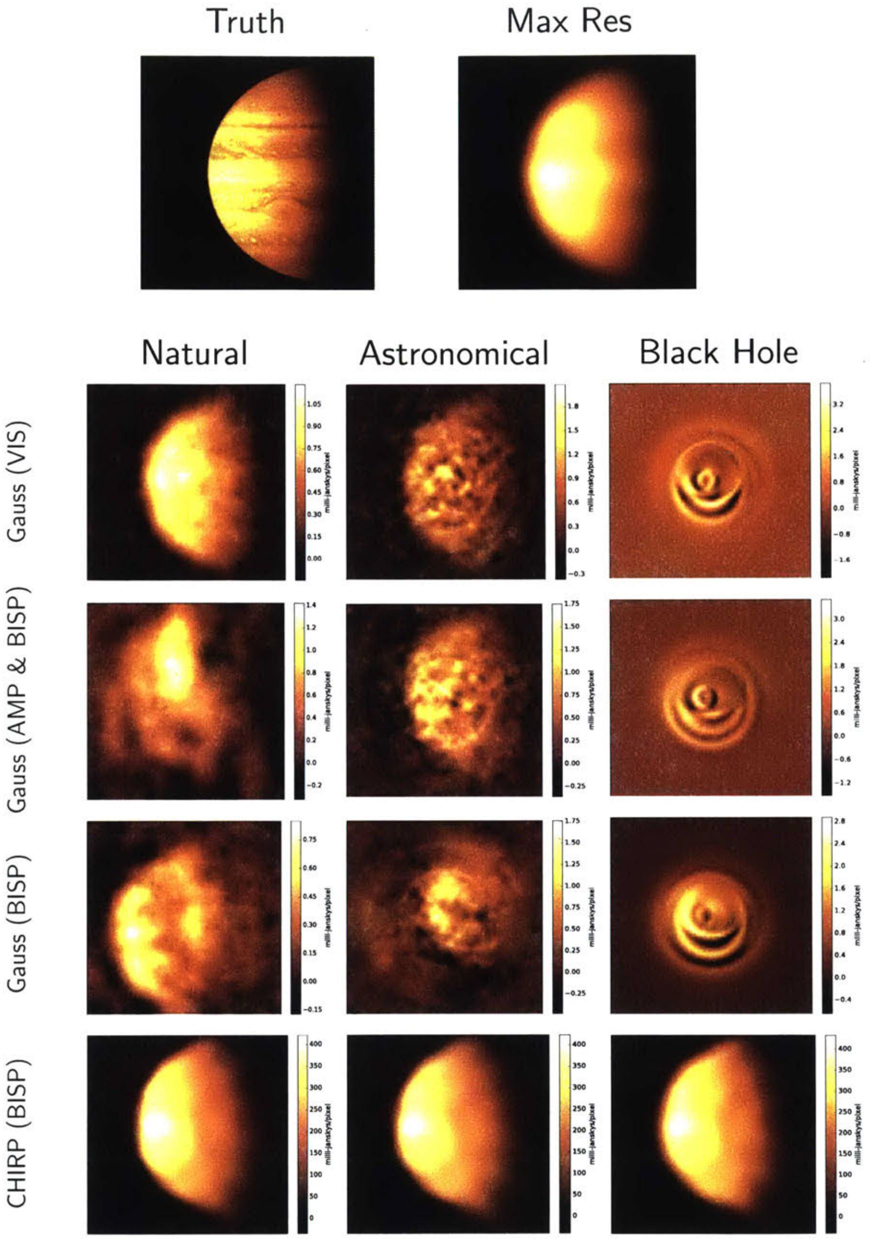


Figure 3.8. Comparing Reconstructions from Trained Models: Reconstructions using image priors (Gauss) and patch priors (CHIRP) trained on natural, celestial, and synthetic black hole images. The truth and blurred truth image are shown on the top. Results of Gauss are shown when constraining the complex visibilities (without atmospheric error), amplitude and bispectrum, as well as just the bispectrum. CHIRP results are only shown when constraining the bispectrum. Notice that significant differences exist between reconstructions when using the different trained image priors, even when clean complex visibilities are provided. However, only small differences can be seen in the reconstructed images using patch priors. Since absolute position is lost during imaging with the bispectrum, shifts in the reconstructed source location are expected.

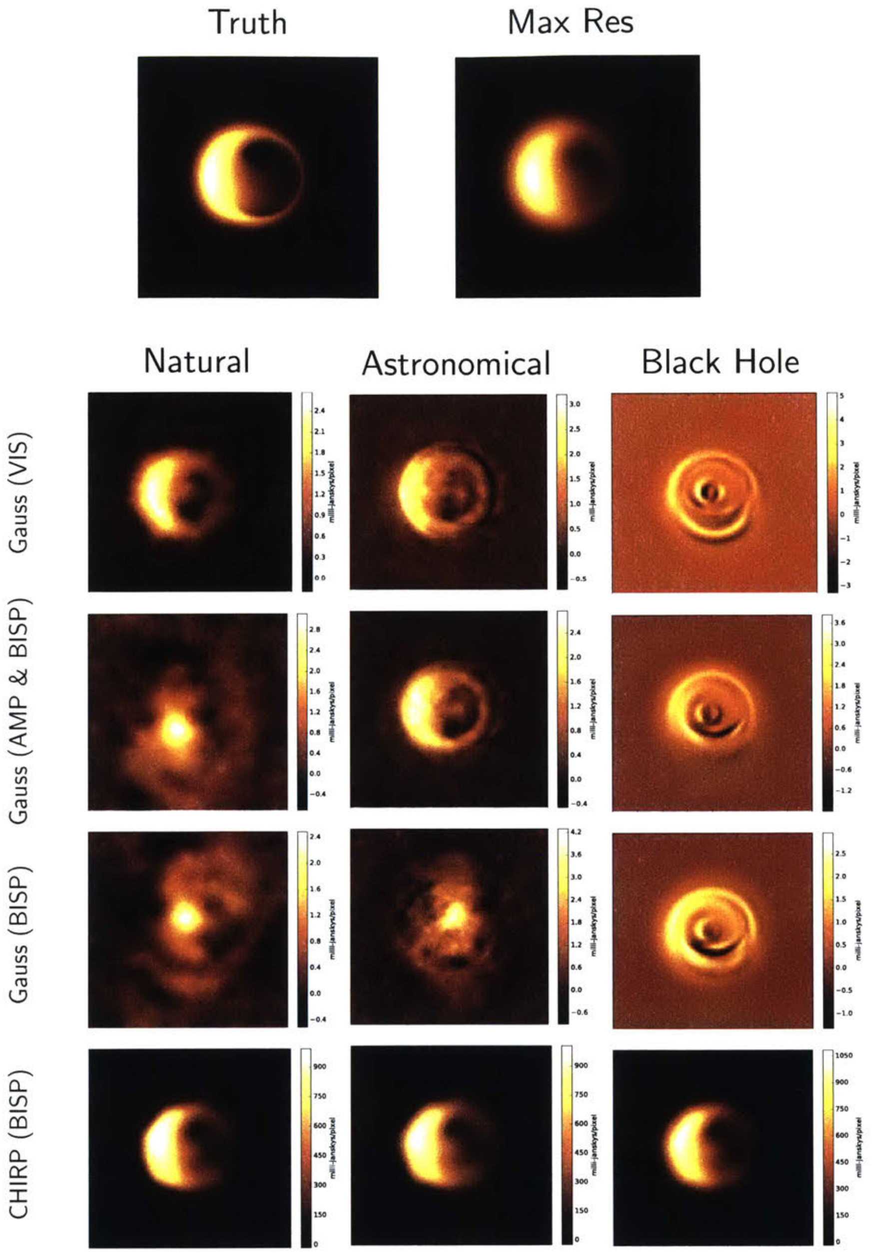


Figure 3.9. Comparing Reconstructions from Trained Models: Refer to the caption of Figure 3.8. In this example the true source image comes from the manifold of images that the 'black hole' prior was trained on. However, despite this the reconstruction using 'Gauss' is very poor. This has to do with the fact that the multivariate Gaussian model is not able to adequately capture the manifold of black hole images. As all of the images used to train the 'black hole' prior are very similar, the resulting model encourages high frequency structure and negative light (see Figure 3.6). This model allows us to easily overfit the data, and find an image that aligns with the noisy data better than the true image.

Patch Model Unlike in the image model, only small variations can be observed among the resulting CHIRP reconstructions done using differently trained patch models. Although each patch model captured different features of the images (see Figure 3.7), CHIRP's reconstruction process is not overly restrictive and is able to select the patches in each model that best describe the observed data. This does not imply that the patch model does not bias image reconstruction at all. In fact, as discussed in Section 5.2.5 of Chapter 5, CHIRP's patch model does bias the appearance of reconstructed images. However, we believe that this experiment implies that the image assumptions that are incorporated into each model do not substantially bias the results; these patch regularizers guide optimization, but do not impose strong assumptions about the overall structure of the underlying image.

■ 3.4.3 Method Comparison

We compare our proposed method, CHIRP, to other state-of-the-art methods on both synthetic and real VLBI data.

Synthetic Data

We compare results of CHIRP (using a natural patch model) with the three state-of-the-art algorithms described in Section 3.1: CLEAN, SQUEEZE, and BSMEM. Images were obtained by asking authors of the competing algorithms or knowledgeable users for a suggested set of reconstruction parameters.

As with our algorithm, SQUEEZE [9] and BSMEM [24] use the bispectrum as input. CLEAN cannot automatically handle large phase errors, so CLEAN results were obtained using *calibrated* (eg. no atmospheric phase error) visibilities in CASA [67]. In reality, these ideal calibrated visibilities would not be available, and the phase would need to be recovered through highly user-dependent self-calibration methods. However, in the interest of a fair comparison, we show the results of CLEAN in a "best-case" scenario.

Figure 3.10 shows a sample of results comparing our reconstructions to those of the current state-of-the-art methods. Our algorithm is able to handle a wide variety of sources, ranging from very simple celestial to complex natural images, without any additional parameter tuning. CLEAN produces consistently blurrier results. Both SQUEEZE and BSMEM tend towards sparser images. This strategy works well for super-resolution. However, it comes at the cost of often making extended sources overly sparse and introducing spurious detail. Although algorithms such as BSMEM and

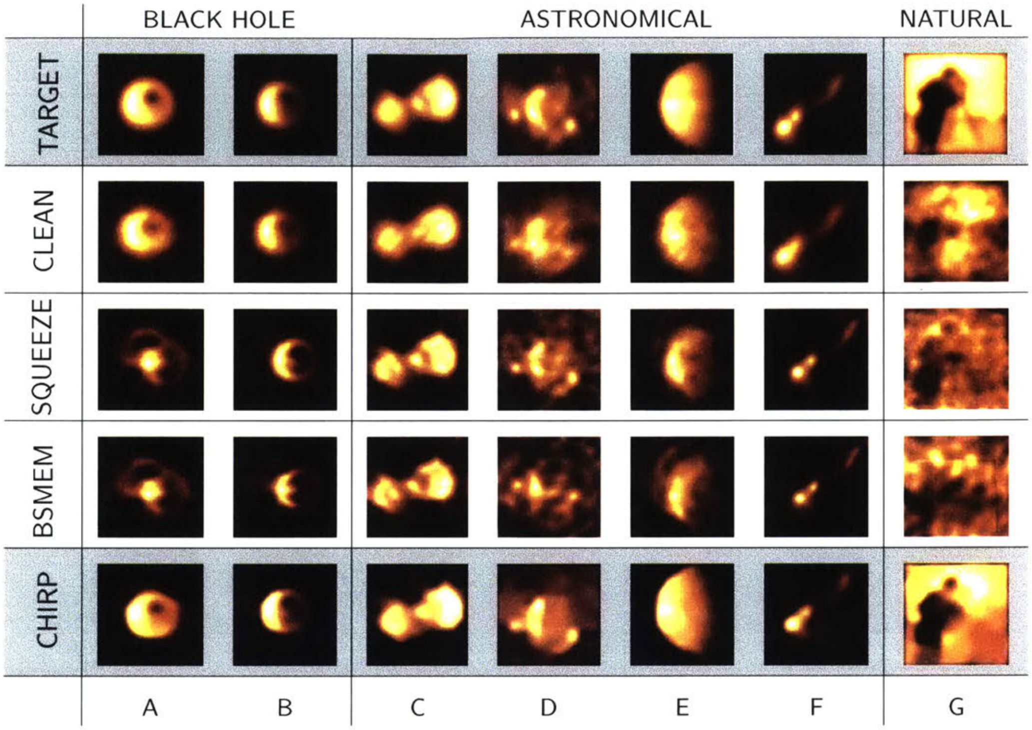


Figure 3.10. Method Comparison: Comparison of our algorithm, 'CHIRP' to three state-of-the-art methods: 'CLEAN', 'SQUEEZE', and 'BSMEM'. We show the normalized reconstruction of a variety of black hole (a-b), astronomical (c-f), and natural (g) source images with a total flux density (sum of pixel intensities) of 1 jansky and a 183.82 μ-arcsecond field of view. Since absolute position is lost when using the bispectrum, shifts in the reconstructed source location are expected. The 'TARGET' image shows the ground truth emission filtered to the maximum resolution intrinsic to this telescope array.

SQUEEZE may perform better on these images with specific hand-tuned parameters, these tests demonstrate that the performance of CHIRP requires less user expertise and provides images that may be less sensitive to user bias.

Figure 3.11 shows a quantitative comparison of our method to SQUEEZE and BSMEM for the blind test set presented in Section 5.1.3 of Chapter 5. Since CLEAN cannot automatically handle large phase errors, we were unable to fairly compare its results on this test set. Since the absolute position of the emission is lost when using the bispectrum, we first align the reconstruction to the ground truth image using cross-correlation. We then evaluate the mean squared error (MSE) and structural similarity index (SSIM) on the normalized, aligned images. Although we consider MSE and SSIM

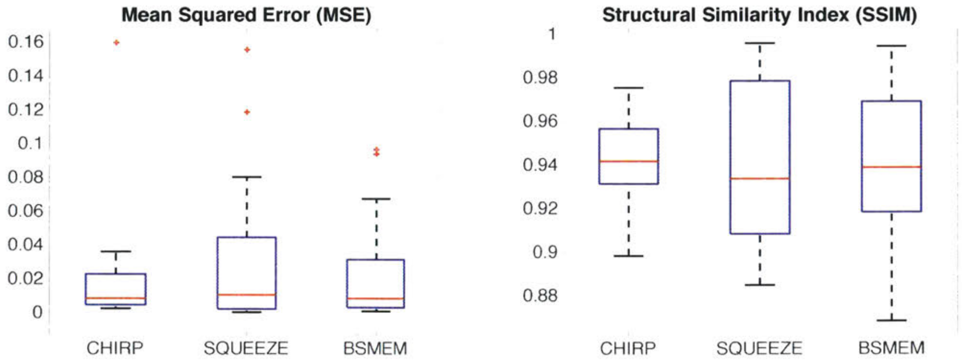


Figure 3.11. Quantitative Analysis on Blind Test Set: Box plots of MSE and SSIM for reconstruction methods on the blind dataset presented in Section 5.1.3. In SSIM a score of 1 implies perceptual indistinguishability between the ground truth and recovered image. Scores are calculated using the original 'Source' image (Refer to Fig. 3.4).

a good first step towards quantitative analysis, we believe a better metric of evaluation is subject for future research.

The thermal noise introduced into each visibility measurement is fixed based upon telescope and measurement choices, and is not dependent on the brightness of the emission source. Consequently, an emission with a lower total flux will result in a lower SNR signal. Previous measurements predict that the total flux densities of the black holes M87 and SgA* will be in the range 0.5 to 3.0 janskys [36, 37]. Figure 3.12 shows the effect of varying total flux density, and thus noise, on each method's recovered reconstructions. Notice how our method is fairly robust to the noise, while the results from other methods often vary substantially across noise levels.

Real Measurements

We demonstrate the performance of our algorithm on the reconstruction of three different sources using real VLBI data from [74] in Figure 3.13. Although we do not have ground truth images corresponding to these measurements, we compare our reconstructions to those generated by the BU group, reconstructed using an interactive calibration procedure described in [74]. Alternatively, we are able to use bispectrum measurements to automatically produce image reconstructions with minimal user input. Notice that we are able to recover sharper images, and even resolve two potentially distinct sources that were previously unresolved in blazar OJ287.

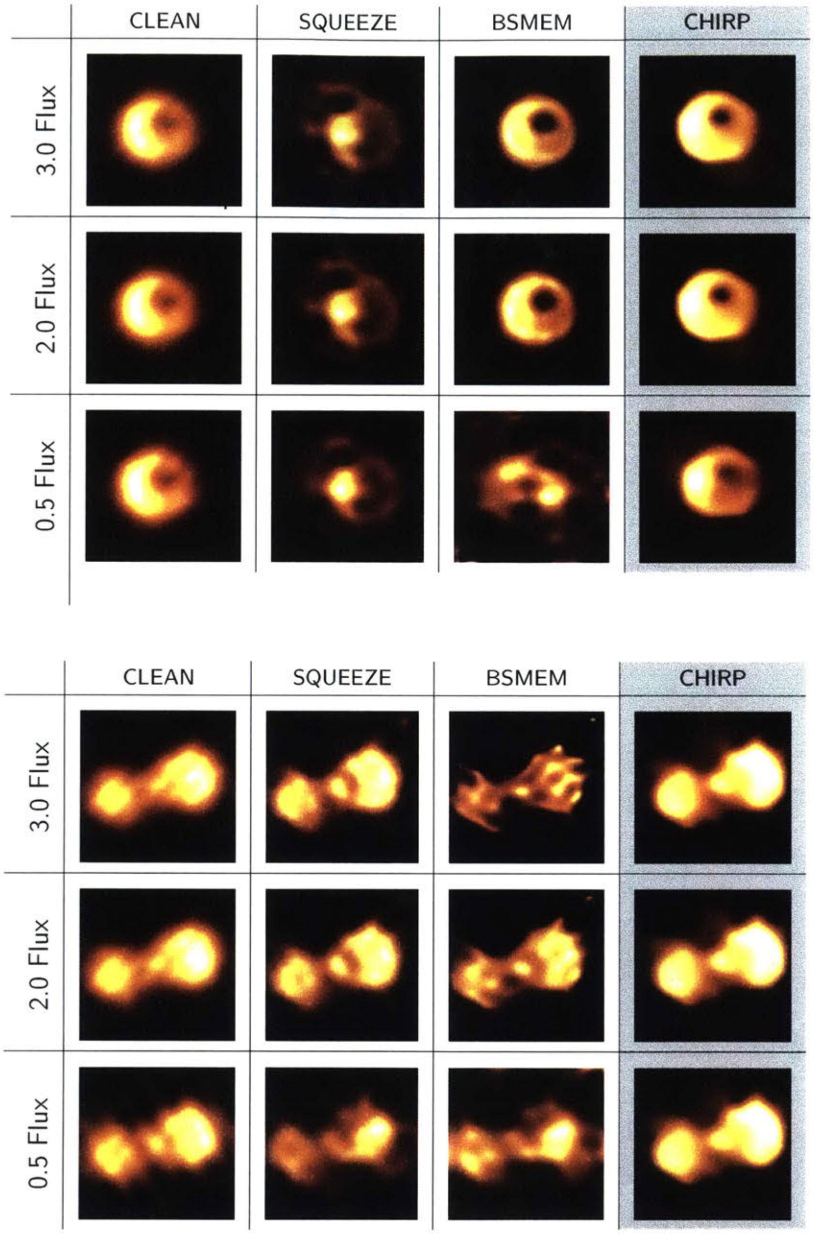


Figure 3.12. Noise Sensitivity: The effect of varying total flux density (in Janskys), and thus noise, on each method's recovered reconstructions. Decreasing flux results in higher noise. Notice how our method is fairly robust to the noise, while the results from other methods often vary substantially across the noise levels. The ground truth target images along with the results for a total flux density of 1 Jansky can been seen in column A and C of Figure 3.10.

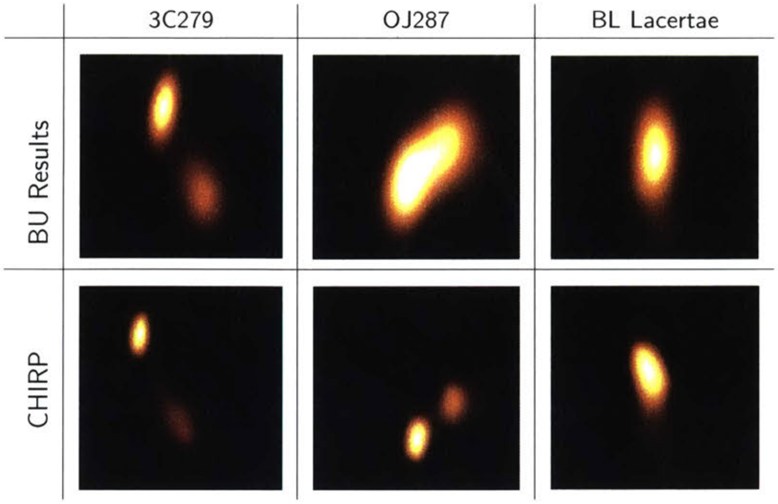


Figure 3.13. Real Measurements: A comparison of our reconstructed images to [74]'s results using CLEAN self-calibration. Note that we are able to reconstruct less blurry images, and are even able to resolve 2 separate, previously unresolved, bright emissions in blazar OJ287. Measurements were taken using the VLBA telescope array. The FOV for each image is 1.5, 1, and 1 milli-arcsecond respectively.

■ 3.5 Conclusion

Astronomical imaging will benefit from the cross-fertilization of ideas with the computer vision community. In this chapter, we have presented an algorithm, CHIRP, for reconstructing an image using a very sparse number of VLBI frequency constraints using data-driven patch priors. This method was developed specifically to handle the challenges faced in the EHT. We have demonstrated improved performance compared to current state-of-the-art methods on both EHT-quality synthetic and real data.

Since CHIRP was initially developed, there have been significant developments in imaging for the EHT. In particular, we incorporated and extended the forward model developed for CHIRP in developing polarimetric imaging methods that use a maximum entropy (MEM) prior [26]. These methods make it possible to study the magnetic field structure around astronomical sources. Additionally, in [6] we explored sparse imaging priors as well as cross-validation techniques to set hyper-parameters during reconstruction. Other recently developed imaging techniques for the EHT can be seen in [5, 42, 69]. When we initially developed CHIRP only the bispectrum was used to constrain reconstruction (as is shown in the results in this chapter). However, after additional analysis it was determined that supplementing this constraint with visibility amplitudes substantially improves performance (see Chapter 5).

4

Video VLBI Reconstruction of Time-Varying Sources

THUS far, very long baseline interferometry (VLBI) has been primarily used to image sources that are static on the time scale of a night's observation. In this chapter we extend the technique's applicability to imaging time-varying sources by reconstructing a video of the source's emission region.

VLBI measurements place a sparse set of constraints on the spatial frequencies of the underlying source image. In particular, each pair of telescopes provides information about a single 2D spatial frequency. This frequency is related to the baseline vector connecting the two telescope sites from the direction of the target source [124]. Thus, at a single time, for an array with P telescopes, at most $P \times (P-1)/2$ spatial frequencies are measured. For example, an array of 6 telescopes would yield only 15 measurements. However, as the Earth rotates, the baseline vector connecting each pair of telescopes changes. This results in sampling additional spatial frequencies along elliptical paths in the frequency plane [124]. Refer to Figure 2.7. Combining the different measurements taken as the Earth rotates is referred to as *Earth Rotation Synthesis.* Earth rotation synthesis is essential for building up enough measurements to constrain image reconstruction.

The task of reconstructing an image from these sparse constraints is highly ill-posed and relies heavily on assumptions made about the underlying image [61, 85, 121]. If a source is static, the VLBI measurements – taken over time as the Earth rotates – all correspond to the same underlying image. Under a static source assumption, recently developed VLBI image reconstruction techniques have been demonstrated on small telescope arrays [5, 6, 17, 26, 42]. However, for an evolving source, measurements are no longer sampled from the same image, and these reconstruction algorithms quickly break down.

Although most astronomical sources are static over the time scale of a night's observation, some notable sources have detectable structural changes on much shorter timescales. For instance, the Galactic Center supermassive black hole, Sagittarius A* (SgrA*), has an estimated mass of only four-million solar masses [50]. This implies that SgrA* is quickly evolving, with an innermost stable circular orbit of just 4 to 30 minutes, depending on the spin of the black hole [72]. Refer to Section 2.4.4 for more detail. Previous observations have shown that SgrA* varies dramatically over a night's observation on the scale of its predicted event horizon, in both total-intensity and polarization [43, 70].

As Earth's closest supermassive black hole, SgrA* is a prime target for the Event Horizon Telescope (EHT) [35]. Unfortunately, the amount of variation predicted for SgrA* suggests that conventional VLBI imaging techniques will be inappropriate for observations taken by the EHT [72]. Thus, in this work we present a new imaging algorithm for time-varying sources that models the VLBI observations as being from a Gaussian Markov Model. Our dynamic imaging algorithm allows for an evolving emission region by simultaneously reconstructing both images and motion trajectories - essentially reconstructing a video rather than a static image.

In this chapter we present our dynamic imaging model for VLBI data and derive an Expectation-Maximization (EM) inference algorithm, StarWarps, that simultaneously reconstructs a movie of the source while also learning the underlying dynamics. In Section 4.1 we discuss how time-variable data differs from static data and also discuss alternate approaches to imaging. Section 4.2 introduces our dynamic imaging model, and in Section 4.3 we discuss how to infer the structure and dynamics of the underlying source using this model. Section 4.4 presents results of our method.

■ 4.1 Alternate Approaches

Earth rotation synthesis inherently assumes that the source being imaged is static over the course of an observation [121]. Generally, this makes it possible to collect multiple measurements that inform us about the same underlying image. However, in the case of an evolving source, as is predicted to be the case for Sgr A*, this assumption is violated – measurements taken at different times throughout the observation correspond to different underlying source images. Thus, VLBI data taken from an evolving source is significantly different from that of a static source [87]. For example, Figure 4.1 shows a simulated visibility amplitude and closure phase expected for a source evolving over

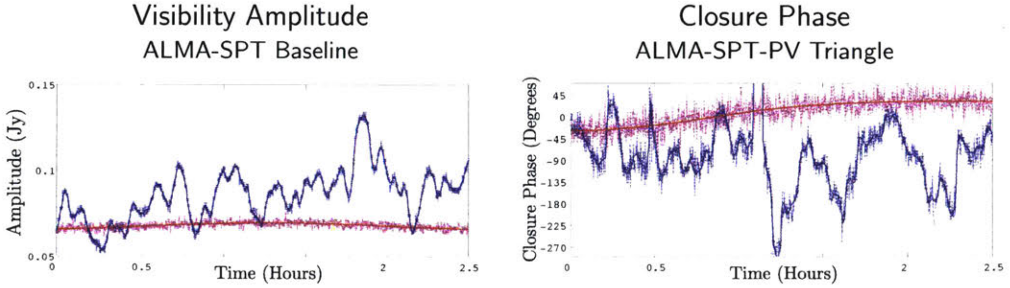


Figure 4.1. Simulated data under a static vs. varying source: Contrasting of data observed from a static emission region (magenta) to that of a varying emission region (blue) over the course of 2.5 hours. Although both sequences start with the same image, the visibility amplitude and closure phase both begin to deviate from the static image very quickly. The ideal observation for the static and time-varying source is shown by the solid red and blue lines, respectively. We also show sample measurements with their respective error bars in the same colors. This data is simulated using the EHT2017 array from the frames in Video 3 presented in Section 4.4.

time, and compares it to the closure phases expected if the same source was static. As the data can no longer be explained by a single image with similar structure, most static imaging methods are unable to produce accurate results.

In recent work the authors attempt to recover a single average image by first smoothing the visibility or bispectrum data taken within a characteristic time-frame [87]. However, as this method was designed to be applied to multi-epoch data, it is often unable to recover an improved reconstruction for a single day's observation when there is significant source variability (see Section 4.4). Additionally, it is not able to recover the dynamics of the underlying source.

In [71] we simultaneously developed an alternative method for time-variability imaging that also solves for a video of the underlying emission region. In that work, we develop a more flexible model with fewer constraints on the imaging assumptions made. However, this modeling choice leads to a much more difficult optimization problem that is prone to local minima, and often requires adhoc methods to achieve satisfactory convergence. The difficulty in optimizing these models can often lead to a solution inconsistent with the true structure of the source images in the case of especially sparse or noisy data. Thus, the strengths and weaknesses of this alternative method are complementary to those of the approach we develop here.

4.2 Dynamic Imaging Model

At each time $t = 1, ..., N$ we measure a vector of data products $\boldsymbol{y}_t$, that are observed from an evolving source image, $\boldsymbol{x}_t$. Our goal is to reconstruct the N instantaneous images $\mathbf{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ using the set of sparse observations $\mathbf{Y} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_N\}$. We define a dynamic imaging model for this observed data as potentials (φ) of an undirected tree graph (see Figure 4.2):

$$\varphi_{\boldsymbol{y}_t|\boldsymbol{x}_t} = \mathcal{N}_{\boldsymbol{y}_t}(f_t(\boldsymbol{x}_t), \boldsymbol{R}_t), \tag{4.1}$$

$$\varphi_{\boldsymbol{x}_t} = \mathcal{N}_{\boldsymbol{x}_1}(\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t), \tag{4.2}$$

$$\varphi_{\boldsymbol{x}_t|\boldsymbol{x}_{t-1}} = \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}), \tag{4.3}$$

for $\boldsymbol{\Lambda}_t = \text{diag}[\boldsymbol{\mu}_t]^T \boldsymbol{\Lambda}' \text{diag}[\boldsymbol{\mu}_t]$.

Similar to the static imaging model, each set of observed data $\boldsymbol{y}_t$ taken at time t is related to the underlying instantaneous source image, $\boldsymbol{x}_t$, through the functional relationship, $f_t(\boldsymbol{x}_t)$, and $\boldsymbol{x}_t$ is encouraged to be a sample from a multivariate Gaussian distribution. However, new to this dynamic imaging model is the addition of (4.3) that describes how images evolve over time. If we assume that there is no evolution between neighboring images in time ($\boldsymbol{A} = \mathbb{1}, \boldsymbol{Q} = \mathbf{0}$), this dynamic model reduces to that of static imaging. Using the Hammersley-Clifford Theorem [55], the joint distribution of this dynamic model can be written as a product of its potential functions:

$$p(\mathbf{X}, \mathbf{Y}|\boldsymbol{A}) \propto \prod_{t=1}^{N} \varphi_{\boldsymbol{y}_t|\boldsymbol{x}_t} \prod_{t=1}^{N} \varphi_{\boldsymbol{x}_t} \prod_{t=2}^{N} \varphi_{\boldsymbol{x}_t|\boldsymbol{x}_{t-1}}. \tag{4.4}$$

4.2.1 Multivariate Gaussian Image Prior

As explained in Chapter 3, a prior distribution on $\boldsymbol{x}$ constrains the the space of possible solutions during inference, and can be defined in a variety of ways. For instance, maximum entropy, sparsity, and patch priors have been all used previously for VLBI imaging [6, 17, 26, 97]. In this work we instead choose to define the underlying image, $\boldsymbol{x}$, as being a sample from the distribution $\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. This choice leads to image reconstructions with reduced sharpness compared to richer priors, but its simplicity allows for a cleaner understanding of our solutions. This proves especially valuable in propagating uncertainties during dynamic imaging (refer to Section 4.3). In Section 3.3.1 we trained a prior of this form using a dataset of images. Although this is possible to do

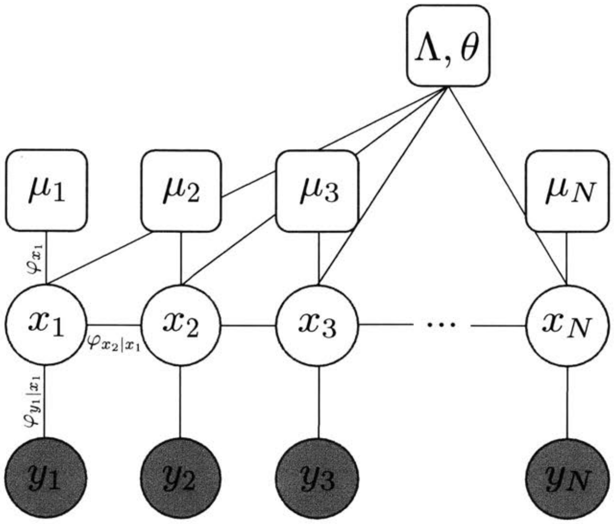


Figure 4.2. Graphical Representation of our Dynamic Imaging Model: At each time t we observe a vector of data products $\boldsymbol{y}_t$ corresponding to the instantaneous source image $\boldsymbol{x}_t$. We assume each image $\boldsymbol{x}_t$ is related to its adjacent neighbors in time, $\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_{t+1}$, and is also related to a multivariate Gaussian distribution specified by mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Lambda}$. The persistent global evolution of the source images over time is specified by $\boldsymbol{A}$, which is further parameterized by θ. Additional intensity perturbations in time are captured by the covariance matrix $\boldsymbol{Q}$. In this diagram, squares indicate parameters, circles are variables, and shaded circles indicate the variable is observed.

here as well, in order to try to avoid the significant bias introduced by these trained models (see Section 3.4.2), we instead hand design the parameters of the prior model according to generic properties that have been widely-used to characterize images.

Image regularizers that enforce spatial smoothness can often be described with a multivariate Gaussian image prior. For instance, the common squared total variation regularizer can be expressed by writing the image covariance, $\boldsymbol{\Lambda}$, in terms of the $2M^4 \times M^4$ gradient matrix, $\boldsymbol{G}$:

$$\boldsymbol{\Lambda} \propto \left[\boldsymbol{G}^T \boldsymbol{G}\right]^{-1}. \tag{4.5}$$

More generally, studies have shown that the average power spectrum of an image often falls with the inverse of spatial frequency in the form $1/(u^2 + v^2)^{\mathsf{a}/2}$, for 2D spatial frequency (u, v) and smoothness parameter a [127]. As the amplitude of a spatial frequency is linearly related to the image itself, this statistical property can also be enforced by specifying the covariance in a prior distribution. Specifically,

$$\boldsymbol{\Lambda}' = \boldsymbol{W}^{*T} \mathrm{diag}\left[\frac{1}{(\boldsymbol{u}^2 + \boldsymbol{v}^2)^{a/2}}\right] \boldsymbol{W} \tag{4.6}$$

for DTFT matrix $\boldsymbol{W}$ of size $M^2 \times M^2$ for an $M \times M$ pixel image. Each row of $\boldsymbol{W}$ corresponds to a (u, v) coordinate in the 2D grid of u and v frequencies, $S \times S$, for

$$S = \left\{\frac{m - M/2}{FOV}\right\}, m \in \mathbb{Z} : m \in [0, M-1], \tag{4.7}$$

where FOV is the image's field of view in radians. To specify the variance of each pixel, and help encourage positivity, we modify the amplitude of the covariance by left and right multiplying by $b \cdot \mathrm{diag}[\boldsymbol{\mu}]$:

$$\boldsymbol{\Lambda} = b^2 \mathrm{diag}[\boldsymbol{\mu}]^T \boldsymbol{\Lambda}' \mathrm{diag}[\boldsymbol{\mu}]. \tag{4.8}$$

A b value of 1/3 implies that 99% of flux values sampled from $\mathcal{N}_x(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ will be positive. Figure 4.4 shows the covariance matrix constructed for $\mathsf{a} = 2, 3, 4$ along with images sampled from the prior $\mathcal{N}_x(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Notice that as a increases, the sampled images are smoother. Thus, a provides the ability to tune the desired smoothness of the inferred images. We compare results of this reconstruction method to other state-of-the-art methods for a static source in Figure 4.3.

■ 4.2.2 Evolution Model

Each image $\boldsymbol{x}_t$ is related to the previous image $\boldsymbol{x}_{t-1}$ through a linear relationship: $\boldsymbol{x}_t \approx \boldsymbol{A}\boldsymbol{x}_{t-1}$. Matrix $\boldsymbol{A}$ (size $M^2 \times M^2$) defines the evolution of the source's emission region between time steps. For instance, $\boldsymbol{A} = \mathbb{1}$ indicates that, on average, the source image does not change, and $\boldsymbol{A} = 2\mathbb{1}$ doubles the image's brightness at each time step. Since the evolution matrix $\boldsymbol{A}$ is not time dependent, the underlying source image evolves similarly over the entire observation.

We assume that the evolution of the emission region over time is primarily described by small perturbations on top of a persistent 2D projected flow of material that preserves total flux. We treat each source image like a 2D array of light pulses originating at locations $(\boldsymbol{\alpha}, \boldsymbol{\delta})$. These pulses can shift around, causing motion in the image. As the position of a pulse can be related to its phase in the frequency domain, under small motions we write $\boldsymbol{A}$ in terms of the image's full DTFT matrix $\boldsymbol{W}$, and a column-vector

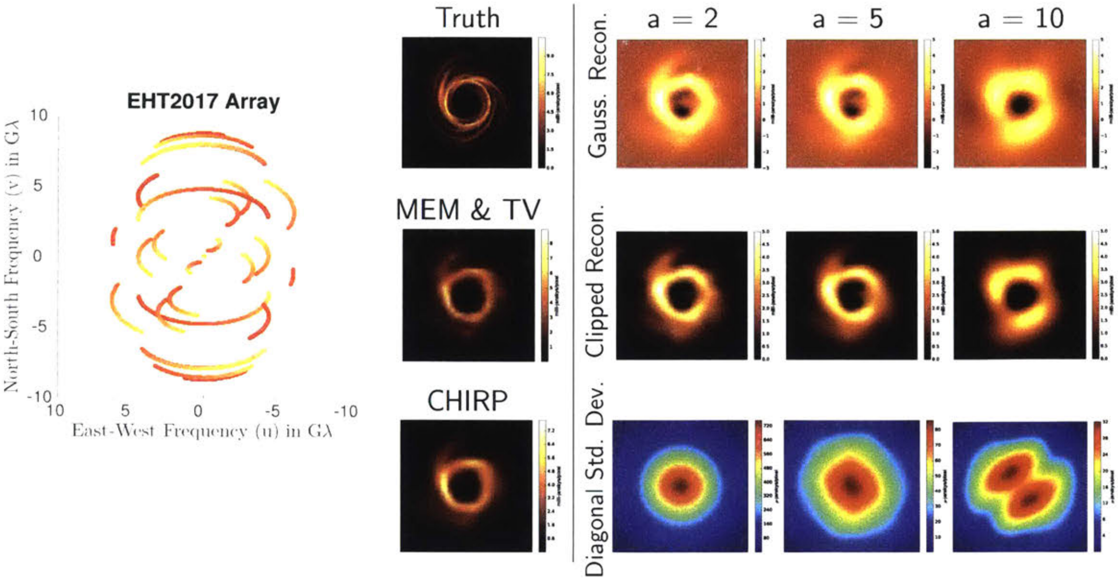


Figure 4.3. Static Imaging Comparison: Results of static imaging using a multivariate Gaussian prior (a = 2, 5, 10) compared to state-of-the-art reconstruction methods using MEM & TV regularizers [26] as well as patch-based regularizers (CHIRP) (see Chapter 3 and [17]). All images are shown with a field of view of 160 μ-arcseconds. Data is generated using a static image with the uv-coverage of the EHT2017 array shown on the left (see Section 4.4). The uv-coverage is colored by time, as indicated by the colorbar in Figure 4.6. Although the previous algorithms (MEM & TV and CHIRP) both produce better results, the Gaussian reconstruction is able to correctly get the broad structure of the underlying image. Since we do not impose positivity, negative values are reconstructed. However, by clipping the resulting image we can see that the result aligns well with the true static image. The Gaussian prior model also allows us to easily estimate our reconstructed image uncertainty. We visualize the diagonal entries of the posterior covariance matrix as the reshaped standard deviation image. Note that as the smoothness parameter a is increased, the per-pixel standard deviation becomes smaller, but the structure of the standard deviation deviates from what was specified in the prior (recall $\boldsymbol{\Lambda}$ is scaled by $\boldsymbol{\mu}$, which we have specified as a 2D Gaussian in this work). For large a the uncertainty is shown to be primarily in the diagonal north-west to south-east direction, due to the lack of spatial frequencies sampled by the telescope array in this direction. To avoid approximations and best show the recovered posterior covariance matrices, atmospheric error has not been included in the data used to recover these images. The flux in Janskys per pixel is normalized across images for a 30 × 30 image. Note we have flipped the uv-coverage relative to how it is normally shown to align with regions that have missing frequency information in the image (in the rest of this thesis we show the uv-coverage plotted as if we were looking towards the Earth rather than towards the source).

of pixel shifts: $\boldsymbol{s} = (\boldsymbol{s}_\alpha, \boldsymbol{s}_\delta)$.

$$\boldsymbol{A} = \Re\left[\boldsymbol{W}^{*T}\boldsymbol{W}\ \exp\left[-i2\pi(\boldsymbol{u}\boldsymbol{s}_\alpha^T + \boldsymbol{v}\boldsymbol{s}_\delta^T)\right]\right]. \tag{4.9}$$

The above parameterization of evolution matrix $\boldsymbol{A}$ in terms of $\boldsymbol{s}$ allows for indepen-

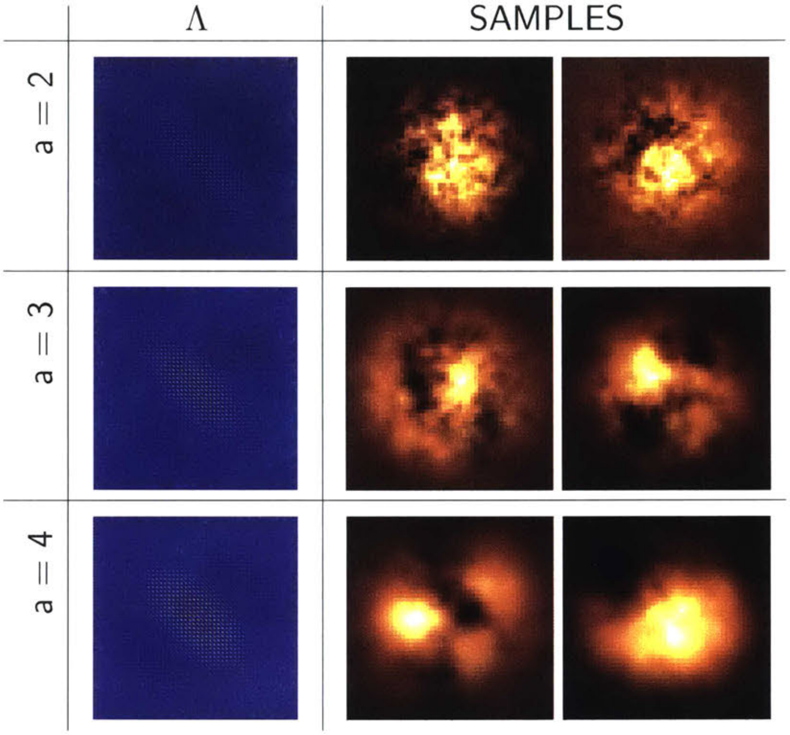


Figure 4.4. Gaussian Image Prior: The covariance matrix constructed for $a = 2, 3, 4$ along with image samples from the prior $\mathcal{N}_x(\mu, \Lambda)$. The image samples have a field of view of 160 μ-arcseconds. Notice that as a increases, the sampled images appear smoother. In these examples μ is a 2D Gaussian Image with standard deviation of 75 μ-arcseconds. and $b = 0.5$. Note that as a increases, the prior encourages smoother structure.

dent, arbitrary shifts of each pulse of light, resulting in $2M^2$ shift parameters. However, as neighboring material generally moves together, the pixel shifts should have a much lower intrinsic dimensionality. To address this, and simultaneously reduce the number of free parameters, we instead describe motion $\boldsymbol{A}$ using a low-dimensional subspace, parameterized by θ. The length of θ, D, is much smaller than the number of unconstrained shift parameters, $2M^2$. We define a motion basis $\mathcal{M} = \left[\mathcal{M}_\alpha, \mathcal{M}_\delta\right]^T$ of size $2M^2 \times D + 1$, and restrict the motion at every time step to be a linear function of this motion:

$$\begin{bmatrix} \boldsymbol{\alpha}_{t+1} \\ \boldsymbol{\delta}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\delta}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{s}_{\alpha_t} \\ \boldsymbol{s}_{\delta_t} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{\alpha_t} \\ \mathcal{M}_{\delta_t} \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix}. \tag{4.10}$$

This parameterization allows us to describe a wide variety of motion (or warp) fields. For instance, affine transformation can be parametrized with a six-dimensional θ that captures rotation, shear, translation, and scaling. More general warp fields can also be described by using a truncated Discrete Cosine Transform (DCT) basis as $\mathcal{M}$.

4.3 Inference & Learning

We solve for the best set of N images $\boldsymbol{X}$ constrained by the N vectors of sparse observations $\boldsymbol{Y}$. In general, we assume that $f_t(.)$, $\boldsymbol{R}_t$, $\boldsymbol{\mu}_t$, $\boldsymbol{\Lambda}_t$, $\boldsymbol{Q}$ are known/specified model parameters. However, $\boldsymbol{A}$, which defines how the source evolves, is not necessarily known ahead of time. If there is reason to believe that only small perturbations exist in the source image over time, then a reasonable assumption is to set $\boldsymbol{A} = \mathbb{1}$. However, in the case of large persistent motion this may fail to give informative results.

We begin in Section 4.3.1 by discussing how to solve for $\boldsymbol{X}$ when $\boldsymbol{A}$ is known. In this case, the model contains no unspecified parameters and the goal is to simply solve for the latent images. In Section 4.3.2, we forgo this assumption and no longer assume that $\boldsymbol{A}$ is known. In this case, we jointly solve for $\boldsymbol{A}$ and $\boldsymbol{X}$ by first learning $\boldsymbol{A}$'s parameters θ using an Expectation-Maximization (EM) algorithm before solving for the latent images, $\boldsymbol{X}$. We refer to our proposed method as StarWarps.

4.3.1 Known Evolution

Given all of the model parameters and observed data, our goal is to estimate the optimal set of latent images, $\boldsymbol{X}$. In static imaging we set up an optimization problem that allowed us to easily solve for the most likely latent image, $\boldsymbol{x}$, given the observed data, $\boldsymbol{y}$. However, in the proposed dynamic model, solving for a MAP estimate – the set of all images that maximizes Equation 4.4 – is much more difficult, as a closed-form solution does not exist. Instead of the MAP estimate, we compute the most likely instantaneous image at each time, t, given all of the observed data $\boldsymbol{Y}$. In particular, we estimate the marginal distribution of each $\boldsymbol{x}_t$, $p(\boldsymbol{x}_t|\boldsymbol{Y})$, by integrating out the other latent images in time, and set $\hat{\boldsymbol{x}}_t$ equal to the mean (and mode) of each distribution.

Since we have defined our dynamic model in terms of Gaussian distributions, we can efficiently solve for $p(\boldsymbol{x}_t|\boldsymbol{y}_1, ..., \boldsymbol{y}_N)$ by marginalizing out the latent images $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t+1}, ..., \boldsymbol{x}_N\}$ using the Elimination Algorithm [80]. Specifically, we derive a function proportional to the marginal distributions. This function is evaluated using a two-pass algorithm, which consists of a forward pass and a backward pass. Each pass, outlined in Algorithms 1 and 3, propagates information using recursive updates that compute distributions proportional to $p(\boldsymbol{x}_t, \boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1})$ and $p(\boldsymbol{y}_t, ..., \boldsymbol{y}_N|\boldsymbol{x}_t)$ for each $\boldsymbol{x}_t$ in the forward and backward pass, respectively. Refer to Appendix B.1 for a derivation of this propagation algorithm. By combining these terms we obtain

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{Y}) &= \mathcal{N}_{x_t}(\hat{\boldsymbol{x}}_t, \boldsymbol{C}_t) \\ &\propto \mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t}), \end{aligned} \tag{4.11}$$

which, as shown in Appendix B.1.3, has mean $\hat{\boldsymbol{x}}_t$ and covariance $\boldsymbol{C}_t$:

$$\begin{aligned} \hat{\boldsymbol{x}}_t &= \mathbf{P}^{\beta}_{t|t}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\alpha}_{t|t-1} + \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\beta}_{t|t} \\ \boldsymbol{C}_t &= \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{P}^{\beta}_{t|t}, \end{aligned} \tag{4.12}$$

where $\mathbf{z}^{\alpha}_{t|\tau}$, $\mathbf{P}^{\alpha}_{t|\tau}$ are the estimates of the mean and covariance of $\boldsymbol{x}_t$ using observations at time steps 1 through τ. Similarly, $\mathbf{z}^{\beta}_{t|\tau}$, $\mathbf{P}^{\beta}_{t|\tau}$ are the estimates of the mean and covariance of $\boldsymbol{x}_t$ using observations τ through N.

For generality, we have listed the forward and backward algorithms in terms of non-linear measurement functions, $f_t(\boldsymbol{x}_t)$ with derivative $\dot{\boldsymbol{F}}$. In this case, similar to our static model inference in Section 3.3.1, we linearize the solution around $\tilde{\boldsymbol{x}}_t$ to get an approximate estimate. To improve the solution of the forward and backward terms, each step in the forward pass can be iteratively re-solved, updating $\tilde{\boldsymbol{x}}_t$ at each iteration. The values of $\tilde{\boldsymbol{x}}_t$ are then fixed for the backwards pass. Recall that when $f_t(\boldsymbol{x})$ is linear in $\boldsymbol{x}$ then $f_t(\boldsymbol{x}) = \boldsymbol{F}_t\boldsymbol{x} = \dot{\boldsymbol{F}}_t\boldsymbol{x}$, and the $\hat{\boldsymbol{x}}$ will converge to the optimal solution in a single update.

The above inference algorithm is similar to the forward-backward algorithm used for Gaussian Hidden Markov Models [80]. In fact, removing the $\varphi_{\boldsymbol{x}_t}$ term for $t > 1$ in Equation 4.4 yields the familiar form of a Gaussian Hidden Markov Model. In this case, inference reduces to the traditional Kalman filtering and smoothing (extended Kalman filtering in the case of non-linear $f_t(\boldsymbol{x})$). Although this simpler formulation can sometimes produce acceptable results, in our typical scenario of especially sparse or noisy data keeping the additional potential terms helps to further constrain the problem, and results in better reconstructions.

Predict:

$$\begin{aligned} \mathbf{z}^{\alpha}_{t|t-1} &= \boldsymbol{A}\mathbf{z}^{\alpha}_{t-1|t-1} \\ \mathbf{P}^{\alpha}_{t|t-1} &= \boldsymbol{Q} + \boldsymbol{A}\mathbf{P}^{\alpha}_{t-1|t-1}\boldsymbol{A}^T \end{aligned}$$

$$\begin{aligned} \mathbf{z}^{\alpha*}_{t|t-1} &= \boldsymbol{\Lambda}_t(\boldsymbol{\Lambda}_t + \mathbf{P}^{\alpha}_{t|t-1})^{-1}\mathbf{z}^{\alpha}_{t|t-1} + \mathbf{P}^{\alpha}_{t|t-1}(\boldsymbol{\Lambda}_t + \mathbf{P}^{\alpha}_{t|t-1})^{-1}\boldsymbol{\mu}_t \\ \mathbf{P}^{\alpha*}_{t|t-1} &= \boldsymbol{\Lambda}_t(\boldsymbol{\Lambda}_t + \mathbf{P}^{\alpha}_{t|t-1})^{-1}\mathbf{P}^{\alpha}_{t|t-1} \end{aligned}$$

Update:

$$\begin{aligned} \boldsymbol{y}_{\Delta} &= (\boldsymbol{y}_t + \dot{\boldsymbol{F}}\tilde{\boldsymbol{x}}_t - f(\tilde{\boldsymbol{x}}_t) - \dot{\boldsymbol{F}}\mathbf{z}^{\alpha*}_{t|t-1}) \\ \mathbf{z}^{\alpha}_{t|t} &= \mathbf{z}^{\alpha*}_{t|t-1} + \mathbf{P}^{\alpha*}_{t|t-1}\dot{\boldsymbol{F}}_t^T(\boldsymbol{R}_t + \dot{\boldsymbol{F}}_t\mathbf{P}^{\alpha*}_{t|t-1}\dot{\boldsymbol{F}}_t^T)^{-1}\boldsymbol{y}_{\Delta} \\ \mathbf{P}^{\alpha}_{t|t} &= \mathbf{P}^{\alpha*}_{t|t-1} - \mathbf{P}^{\alpha*}_{t|t-1}\dot{\boldsymbol{F}}_t^T(\boldsymbol{R}_t + \dot{\boldsymbol{F}}_t\mathbf{P}^{\alpha*}_{t|t-1}\dot{\boldsymbol{F}}_t^T)^{-1}\dot{\boldsymbol{F}}_t\mathbf{P}^{\alpha*}_{t|t-1} \end{aligned}$$

Initialization:

$$\mathbf{z}^{\alpha*}_{1|0} = \boldsymbol{\mu}_1 \ , \ \mathbf{P}^{\alpha}_{1,0} = \boldsymbol{\Lambda}_1$$

Algorithm 1. Forward Updates $t = 1 \rightarrow 2 \rightarrow \ldots \rightarrow N$

4.3.2 Unknown Evolution

If the evolution matrix $\boldsymbol{A}$ is unknown, we are unable to solve for $\boldsymbol{X}$ in the way outlined in Section 4.3.1. In theory, although it is possible to compute a gradient of Equation 4.4 with respect to each $\boldsymbol{X}$ and $\boldsymbol{A}$ and perform gradient ascent, this would be very computationally intensive and prone to local maxima. This becomes even more pronouced when solving for the evolution parameters θ (see Section 4.2.2) rather than $\boldsymbol{A}$ itself.

Instead we choose to jointly solve for latent images $\boldsymbol{X}$ and parameters $\boldsymbol{A}$ (parameterized by θ) using the Expectation-Maximization (EM) algorithm. The EM algorithm defines an iterative process that solves for the evolution parameters θ that maximize the complete likelihood in Equation 4.4 when all of the underlying images, $\boldsymbol{X}$, are unknown (latent). Each iteration of EM improves the log-likelihood of the data under the defined objective function and is especially useful when the likelihood is from an exponential family, as is the case in our proposed model. In particular the EM algorithm consists of the following two iterative steps:

Predict:

$$\mathbf{z}^{\beta*}_{t|t+1} = \boldsymbol{\mu}_t + \boldsymbol{\Lambda}_t \boldsymbol{A}^T (\boldsymbol{Q} + \mathbf{P}^{\beta}_{t+1|t+1} + \boldsymbol{A}\boldsymbol{\Lambda}_t \boldsymbol{A}^T)^{-1} (\mathbf{z}^{\beta}_{t+1|t+1} - \boldsymbol{A}\boldsymbol{\mu}_t)$$
$$\mathbf{P}^{\beta*}_{t|t+1} = \boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t \boldsymbol{A}^T (\boldsymbol{Q} + \mathbf{P}^{\beta}_{t+1|t+1} + \boldsymbol{A}\boldsymbol{\Lambda}_t \boldsymbol{A}^T)^{-1} \boldsymbol{A}\boldsymbol{\Lambda}_t$$

Update:

$$\boldsymbol{y}_{\Delta} = (\boldsymbol{y}_t + \dot{\boldsymbol{F}}\tilde{\boldsymbol{x}}_t - f(\tilde{\boldsymbol{x}}_t) - \dot{\boldsymbol{F}}_t \mathbf{z}^{\beta*}_{t|t+1})$$
$$\mathbf{z}^{\beta}_{t|t} = \mathbf{z}^{\beta*}_{t|t+1} + \mathbf{P}^{\beta*}_{t|t+1} \dot{\boldsymbol{F}}_t^T (\boldsymbol{R}_t + \dot{\boldsymbol{F}}_t \mathbf{P}^{\beta*}_{t|t+1} \dot{\boldsymbol{F}}_t^T)^{-1} \boldsymbol{y}_{\Delta}$$
$$\mathbf{P}^{\beta}_{t|t} = \mathbf{P}^{\beta*}_{t|t+1} - \mathbf{P}^{\beta*}_{t|t+1} \dot{\boldsymbol{F}}_t^T (\boldsymbol{R}_t + \dot{\boldsymbol{F}}_t \mathbf{P}^{\beta*}_{t|t+1} \dot{\boldsymbol{F}}_t^T)^{-1} \dot{\boldsymbol{F}}_t \mathbf{P}^{\beta*}_{t|t+1}$$

Initialization:

$$\mathbf{z}^{\beta*}_{N|N+1} = \boldsymbol{\mu}_N \ , \ \mathbf{P}^{\beta*}_{N|N+1} = \boldsymbol{\Lambda}_N$$

Algorithm 2. Backward Updates: $t = N \rightarrow N-1 \rightarrow \ldots \rightarrow 1$

- Expectation step (E step): Calculate the expected value of the log likelihood function, with respect to the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ under the current estimate of the θ parameters, $\theta^{(i)}$:

$$Q(\theta|\theta^{(i)}) = E_{\mathbf{X}|\mathbf{Y},\theta^{(i)}}[\log p(\mathbf{X},\mathbf{Y}|\theta)] \tag{4.13}$$

- Maximization step (M step): Find the parameter that maximizes this quantity:

$$\theta^{(i+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(i)}) \tag{4.14}$$

We solve for the optimal θ using gradient ascent. As $\boldsymbol{A}$ is a function of θ, we must compute the derivative of $Q(\theta|\theta^{(i)})$ using the chain rule. We compute this derivative with respect to each element j in θ:

$$\frac{d}{d\theta[j]} Q(\theta|\theta^{(i)}) = \sum_p \sum_q \frac{dQ(\theta|\theta^{(i)})}{d\boldsymbol{A}[p,q]} \frac{d\boldsymbol{A}[p,q]}{d\theta[j]} \tag{4.15}$$

Using the low-dimensional subspace evolution model proposed in Section 4.2.2, the derivative $\frac{d\boldsymbol{A}[p,q]}{d\theta[j]}$ can be computed as

$$\frac{d\boldsymbol{A}}{d\theta[j]} = -i2\pi\theta[j]\boldsymbol{A}\left(\boldsymbol{u}\mathcal{M}_{\alpha}[:,j+1]^T + \boldsymbol{v}\mathcal{M}_{\delta}[:,j+1]^T\right). \tag{4.16}$$

By expanding and taking the derivative of the log-likelihood from Equation 4.4 with respect to $\boldsymbol{A}$, we obtain the expression

$$\begin{aligned}\frac{d}{d\boldsymbol{A}}Q(\theta|\theta^{(i)}) = -\frac{1}{2}\sum_{t=2}^{N}\Big[&2\boldsymbol{Q}^{-1}\boldsymbol{A}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^T\right] \\ &-\boldsymbol{Q}^{-1}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_{t-1}^T\right] - \boldsymbol{Q}^{-1}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right]\Big]\end{aligned} \tag{4.17}$$

By inspecting this expression we can see that the sufficient statistics we require to maximize the log-likelihood are the expected value of $\boldsymbol{x}_t\boldsymbol{x}_t^T$ and $\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T$ under the distribution $p(\boldsymbol{X}|\boldsymbol{Y};\theta^{(i)})$. From the marginal distributions (Equations 4.11 and 4.12) derived in Section 4.3.1, we obtain

$$E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_t^T\right] = \mathbf{z}_t\mathbf{z}_t^T + \mathbf{P}_t. \tag{4.18}$$

The sufficient statistic $E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right]$ is a bit trickier to obtain, but can also be calculated using the same forward-backward terms, as shown in Appendix B.2. Mathematically,

$$E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right] = \mathbf{z}_{t-1}\mathbf{z}_t^T + \boldsymbol{\xi}_3\boldsymbol{\xi}_1^{T-1}, \tag{4.19}$$

where

$$\begin{aligned}p(\boldsymbol{x}_t,\boldsymbol{x}_{t-1}|\boldsymbol{Y}) &= \mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{\xi}_1\boldsymbol{x}_t + \xi_2, \boldsymbol{\xi}_3) \\ &\propto \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}_{t-1|t-1}^{\alpha}, \mathbf{P}_{t-1|t-1}^{\alpha})\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}_{t|t}^{\beta}, \mathbf{P}_{t|t}^{\beta})\end{aligned} \tag{4.20}$$

To learn the parameters θ, we iterate between computing sufficient statistics of $\boldsymbol{X}$ given the current estimate of parameters, $\theta^{(i)}$, and solving for new parameters that maximize the updated log-likelihood, $\theta^{(i+1)}$, under those statistics. Once the algorithm converges, we use the resulting $\hat{\theta}$ to compute $\boldsymbol{A}$ and infer the best set of instantaneous images, $\boldsymbol{X}$ using the two-pass method described in Section 4.3.1.

4.4 Results

4.4.1 Data Generation

We demonstrate our algorithm on synthetic data generated from four different sequences of time-varying sources. These sequences include two realistic GRMHD models of a black hole at different orientations [118], a realistic sequence of a "hot spot" rotating around a black hole [21], and a toy sequence evolving with pure rotation. The field of view of each sequence ranges from 120 to 160 μ-arcseconds. A still frame from each sequence is shown in Figure 4.5. To help give a sense of the variation in each sequence, the figure also displays the mean and standard deviation of flux density. We refer to these sequences by their video number, indicated in the figure.

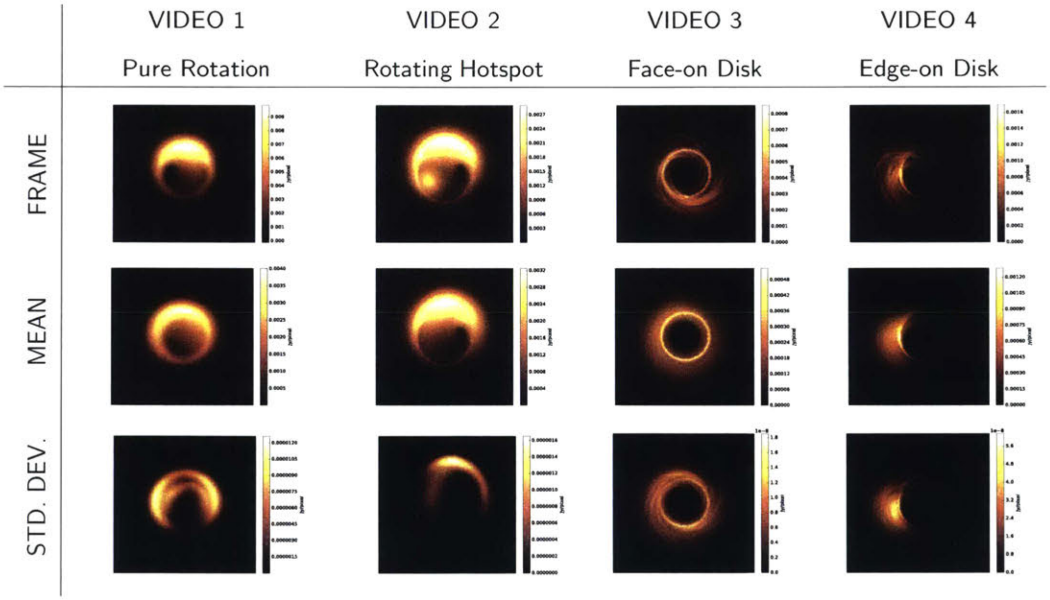


Figure 4.5. Ground truth videos: The four ground truth sequences used to demonstrate results. We show a single frame from each sequence, the mean frame, and the spatial standard deviation of flux density. Video 1 consists of a 160μ-arcsecond image [23] that rotates 180° over the course of a 12 hour observation (24 hour rotational period). Video 2 is a 120μ-arcsecond view of an edge-on black hole disk with a rotating "hot spot" predicted by [21] with a rotational period of 2.78 hours. Video 3 and 4 are generated using a GRMHD model of a black hole observed face on and at a 45° inclination with a 160μ-arcsecond field of view [118]. They assume a spin of 0.9375 with an Innermost Stable Circular Orbit (ISCO) rotational period of 8.96 minutes.

In order to demonstrate the quality of results under various observing conditions, VLBI observations of SgrA* at 1.3 mm (230 GHz) are simulated assuming three different telescope arrays. The first array, EHT2017, consists of the 8 telescopes at 6 distinct locations that were used to collect measurements for the Event Horizon Telescope in the spring of 2017. The second array, EHT2017+, augments the EHT2017 array with 3 additional telescopes that are soon to be added or being explored as potential additions to the EHT: Plateau de Bure (PDB), Haystack (HAY), and Kitt Peak (KP) Observatory. Details on telescopes used in the EHT2017 and EHT2017+ array are shown in Table 2.2. The third array, FUTURE, consists of 9 additional telescopes. The uv-coverage of these arrays, along with a colorbar indicating the time of each measurement, is shown in Figures 4.6.

Visibility measurements are generated using the python `eht-imaging`[1] library [26]. Realistic thermal noise, resulting from a bandwidth ($\Delta\nu$) of 4 GHz and a 100 second integration time (t_{int}), is introduced on each visibility. The standard deviation of thermal noise is given by

$$\sigma = \frac{1}{0.88}\sqrt{\frac{\text{SEFD}_1 \times \text{SEFD}_2}{2 \times \Delta\nu \times t_{\text{int}}}}, \tag{4.21}$$

for System Equivalent Flux Density (SEFD) of the two telescopes corresponding to each visibility[2] [121]. Atmospheric phase error is also introduced into measurements using the eht-imaging library. In Videos 2-4 a set of measurements is sampled every 5 minutes over a roughly 14 hour duration, resulting in 173 time steps. In Video 1 only 30 time steps are measured over a 12 hour duration.

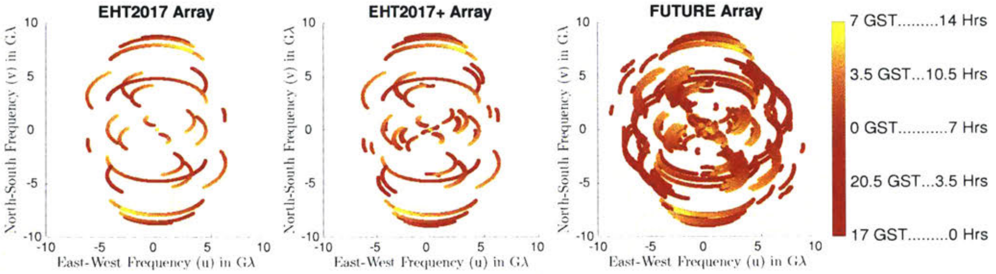


Figure 4.6. Time-varying uv-coverage: The uv-coverage for EHT2017, EHT2017+ and FUTURE arrays when observing SgrA. Baselines are colored by the time of each observation relative to the start time, indicated by the colorbar to the right.

---

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

[2] The factor of 1/0.88 is due to information loss due to recording 2-bit quantized data-streams at each telescope [124].

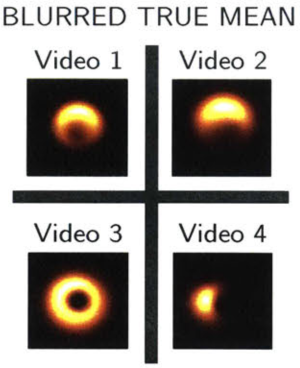
4.4.2 Static Evolution Model (No Warp)

We first demonstrate results of our method under a static evolution model. In this case, we fix parameters θ such that $A = \mathbb{1}$. This assumes that there is no global motion under a persistent warp field, but only perturbations around a fairly static scene. Despite this incorrect assumption (especially in Videos 1 and 2), this simple model results in reconstructions that surpass the state-of-the-art methods, and recovers distinctive structures that appear in the underlying source images.

Figure 4.7 shows example reconstructions, and corresponding measured error (NRMSE), for combinations of the 4 source videos observed under the 3 telescope arrays. The main portion of this figure is broken up into 4 quadrants, each containing results for one video. From left to right, up to down, each quadrant corresponds to Video 1-4 respectively. The ground truth mean image for each video is shown in the upper table. These images correspond to those shown in Figure 4.5, but are smoothed to 3/4 the nominal resolution of the interferometer to help illustrate the level of resolution we aim to recover.

Horizontally We present results obtained using data with varying degrees of difficulty. As the number of telescopes in the array increases, so does the spatial frequency coverage. Therefore, reconstructing an accurate video with the FUTURE array is a much easier task than with the EHT2017 array. Additionally, using complex visibilities that are not subject to atmospheric errors is much easier than having to recover images from phase corrupted measurements. In the case where there are atmospheric phase errors (ATM.), we constrain the reconstruction problem using a combination of visibility amplitude and bispectrum data products. This results in a non-convex problem (that we approximate with series of linearizations) that is much more difficult to solve than when using complex visibilities when there is no atmospheric phase error (NO ATM.). We demonstrate results on the EHT2017 array for both cases, and the EHT2017+ and FUTURE arrays in the case of atmospheric error.

Vertically Within each quadrant we illustrate the results of our method, StarWarps, by displaying the average frame reconstructed. We compare our method to two state-of-the-art Bayesian-style methods. [26] solves for a single image by imposing a combination of MEM and TV priors. This method performs well in the case of a static source (see Figure 4.3), however, in the case of an evolving source it often results in artifact-heavy reconstructions that are difficult to interpret. In [87] the authors attempt to mitigate



NORMALIZED RMSE

	EHT 2017 NO ATM.	EHT 2017 ATM.	EHT 2017+ ATM.	FUTURE ATM.	EHT 2017 NO ATM.	EHT 2017 ATM.	EHT 2017+ ATM.	FUTURE ATM.
StarWarps Mean	**0.67**	**0.65**	**0.55**	**0.55**	**0.74**	**0.73**	**0.73**	0.71
[87]	1.05	0.82	0.73	0.68	0.98	1.21	0.99	0.35
[26]	0.79	0.80	0.73	0.63	0.83	1.05	0.81	**0.23**
StarWarps Mean	**0.32**	**0.55**	0.67	**0.36**	**0.34**	**0.44**	**0.23**	**0.31**
[87]	0.93	0.90	**0.53**	0.51	0.84	0.75	1.02	0.61
[26]	0.84	0.85	0.71	0.39	0.60	0.50	0.60	0.47

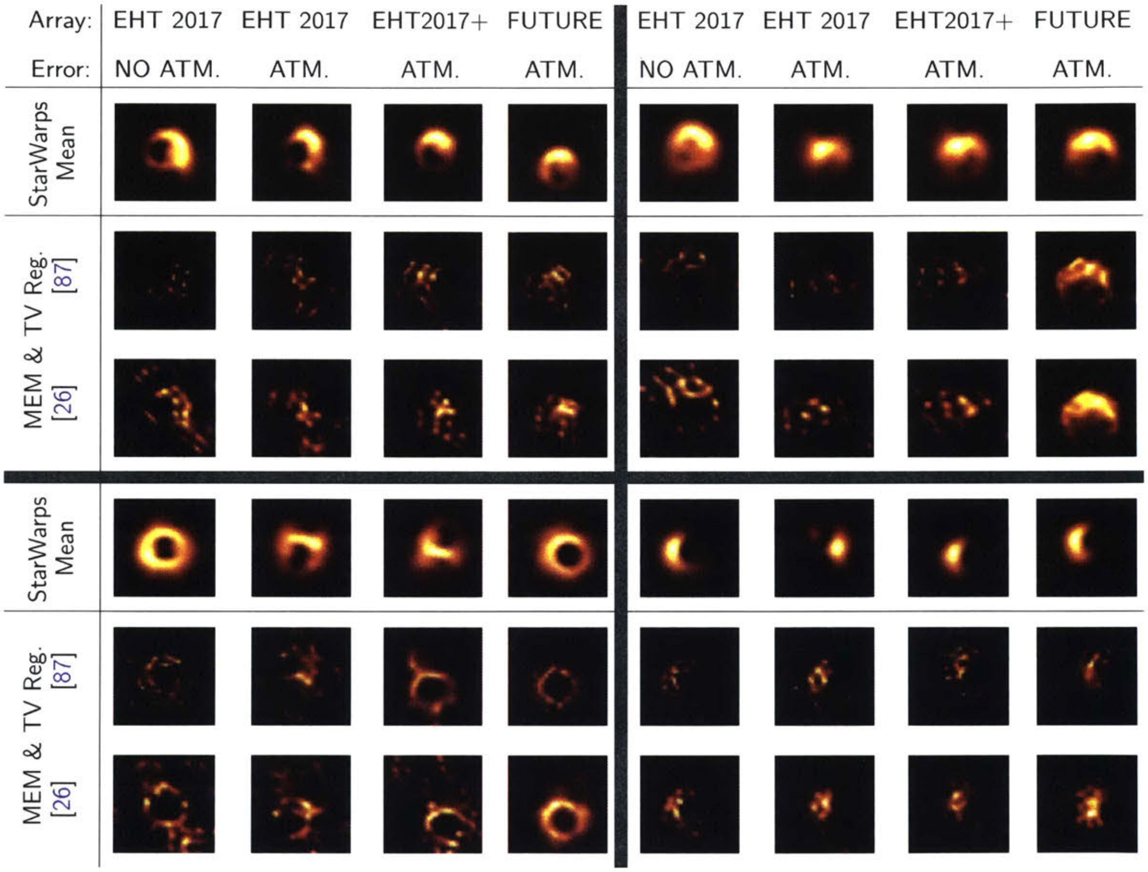


Figure 4.7. Static evolution model: Results obtained using data simulated from each of the 4 video sequences (see Figure 4.5) under different telescope arrays (see Figure 4.6) and noise conditions. The main portion of the figure is broken up into 4 quadrants corresponding to Videos 1-4 when moving from left to right, top to bottom. The true mean image from the ground truth videos, blurred to 3/4 the nominal resolution of the array, is shown on the top. We compare results of our proposed method, StarWarps, to that of the single imaging methods presented in [87] and [26]. In particular, we compare the mean image obtained using StarWarps video reconstruction. The error type NO ATM. indicates reconstructing using visibilities on data with no atmospheric error, while the error type ATM. indicates using the visibility amplitudes and bispectrum on data where atmospheric phase errors have been introduced. The quality of each result, compared to the ground truth mean image, is indicated in the table of normalized root mean squared errors (Normalized RMSE). To account for the loss of absolute position in the presence of atmospheric phase error, images were rigidly aligned to the true mean before computing the error.

this problem by first smoothing the time-varying data products before imaging. This approach was originally designed to work on mutli-epoch data; we find it is unable to accurately recover the source structure from a single day (epoch) observation. Results of [87] are reconstructed by an author of the method.

■ 4.4.3 Unknown Evolution Model (Learn Warp)

In Section 4.4.2 we showed that a static model can often substantially improve results over the state-of-the-art methods, even when there is significant global motion. However, when a source's emission region evolves in a similar way over time, we are able to further improve results by simultaneously estimating a persistent warp field along with the video frames. We demonstrate the StarWarps EM approach proposed in Section 4.3.2, on Videos 1 and 2. In results presented, we have assumed an affine motion basis with no translation (θ consists of 4 parameters), and have allowed the method to converge over 30 EM iterations.

Figure 4.8 shows the recovered warp field recovered by our EM algorithm. Results were obtained from data with and without atmospheric error. In Video 1 the true underlying motion of the emission region can be perfectly captured by the affine model we assume. This allows us to freely recover a very similar warp field. However, in the "hot spot" video (Video 2), although this is no longer the case, we still recover an accurate estimate indicating the direction of motion.

Figures 4.9 and 4.10 compare results obtained when we assume no global motion ($A = \mathbb{1}$) to those when we allow the method to search for a persistent warp field. Results are shown in two settings: when data is generated using the EHT2017+ array assuming no atmospheric phase error, as well as when phase errors are introduced into the measurements. At each time, only a small number of measurements are observed (indicated by the corresponding uv-coverage). However, by propagating information across the video we are able to reconstruct good quality images at each time step. In the case of large global motion, most of the reconstructed motion is suppressed when we assume $A = \mathbb{1}$. However, by solving for the low dimensional parameters of the warp field, θ, we can learn about the underlying dynamics and sometimes produce higher quality videos. Figure 4.11 helps to further visualize the recovered motion in the "hot spot" video by showing how the intensities of a region evolve over time.

Results of our method are compared to that of a simple baseline method that we refer to as 'snapshot imaging'. In snapshot imaging each frame of the video is independently reconstructed using only the small number of measurements taken at that time step.

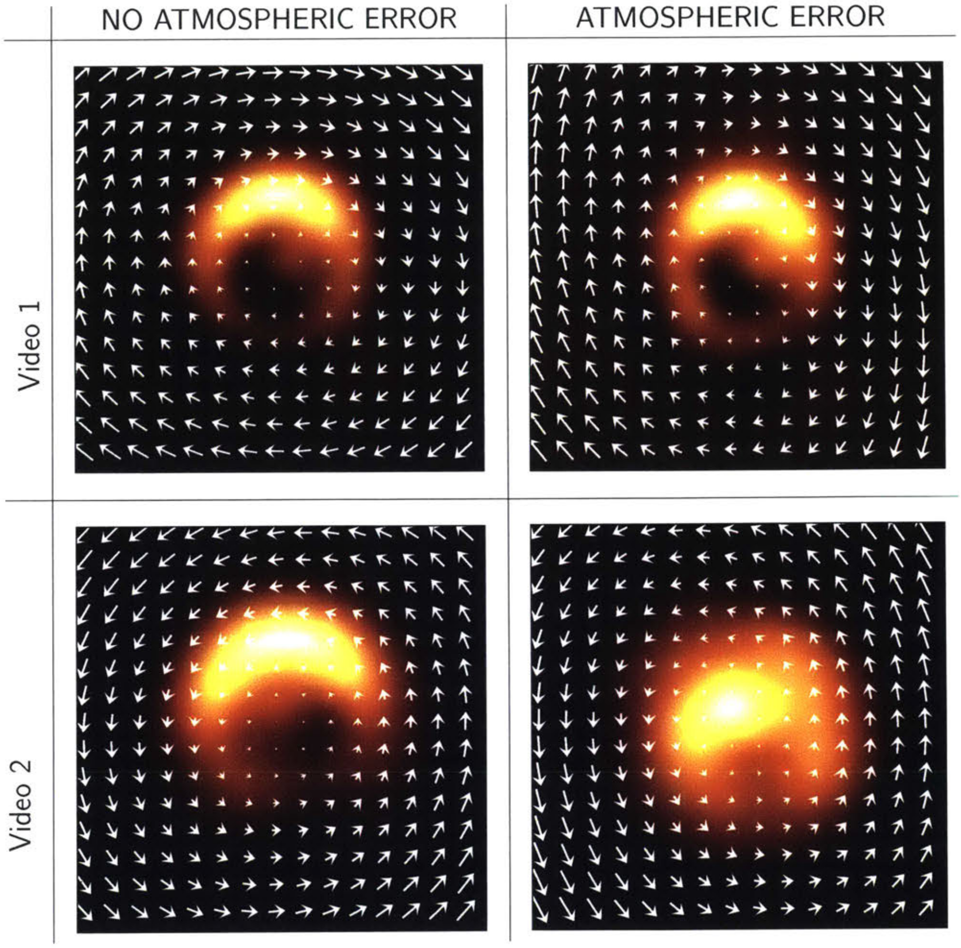


Figure 4.8. Recovering Warp Field: By solving for the parameters of a persistent warp field using the proposed EM algorithm, we are able to recover a low-dimensional representation of the source dynamics. Results are shown using the EHT2017+ array with and without atmospheric error (ATMOSPHERIC and NO ATMOSPHERIC ERROR, respectively). Arrows showing the direction of recovered motion are overlaid on the mean image for a recovered video. In Video 1 the true underlying motion can be described by a clockwise rotation. The proposed method is able to recover Video 1's motion from the observed data. Video 2 contains a 'hot spot' rotating counter-clockwise around a static emission. Video 2 cannot be described using a single persistent flow field. Yet, despite this, the proposed method is still able to recover the general direction of counter-clockwise motion.

In particular, we use the MEM & TV method shown in Figure 4.3 to reconstruct each snapshot [26]. In the case of using complex visibilities, both our method and snapshot imaging produces meaningful results. Although distinctive features of the true underlying image are recovered by both methods, the quality of our StarWarps

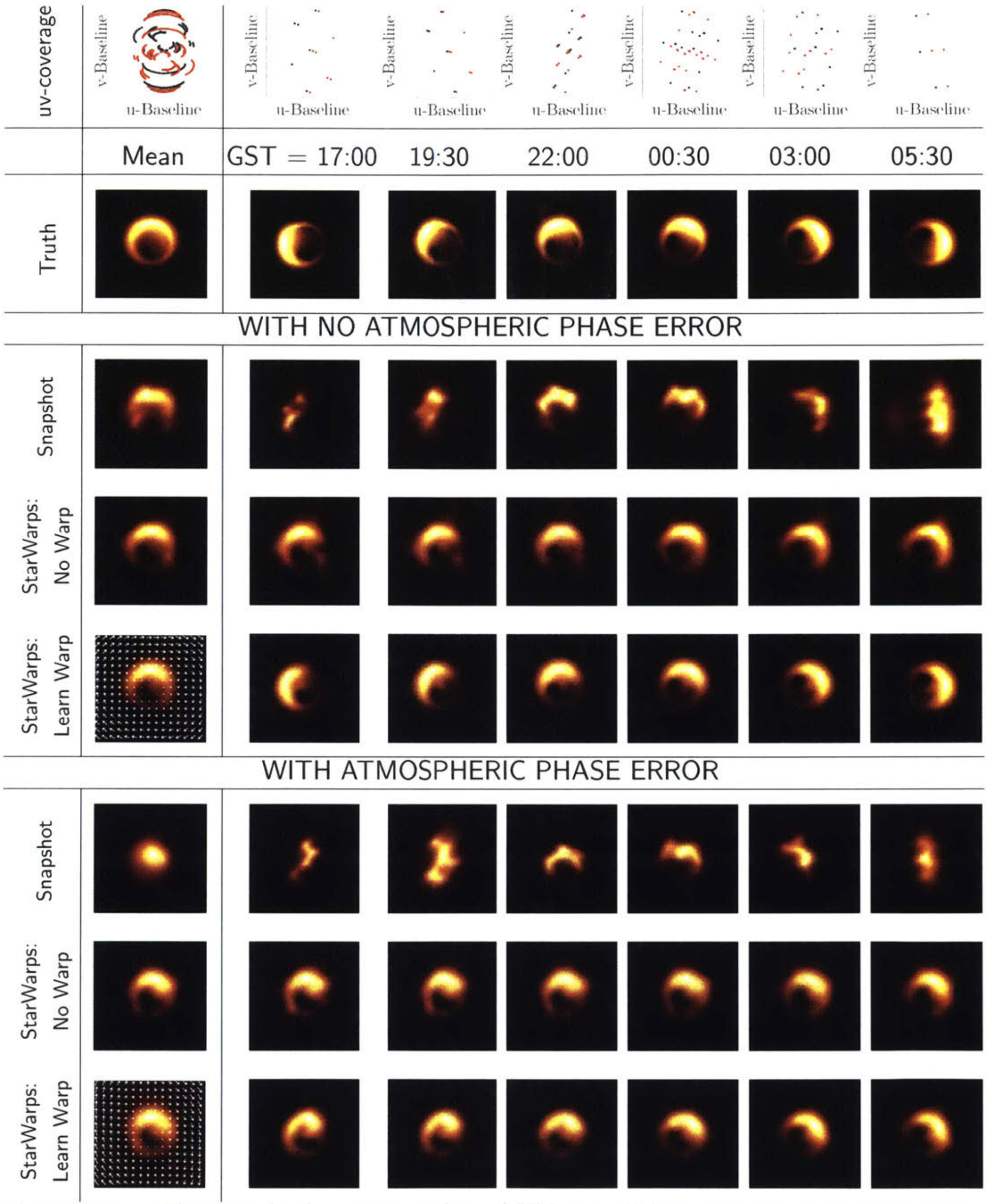


Figure 4.9. **Time-resolved reconstruction of Video 1:** Video 1 contains an image rotating clockwise by 180° over the course of the observation. At each time, the interferometric telescope array measures values related to 2D spatial frequencies of the current underlying image, shown in the row labeled 'Truth'. These are indicated by the dots on the uv-coverage plots (each independent set of measurements displayed as either black or red). We present results obtained when using calibrated data with no atmospheric error, as well as when there is atmospheric phase error still present and we must use data products invariant to its effects. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained when we assume a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. In the case that we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image.

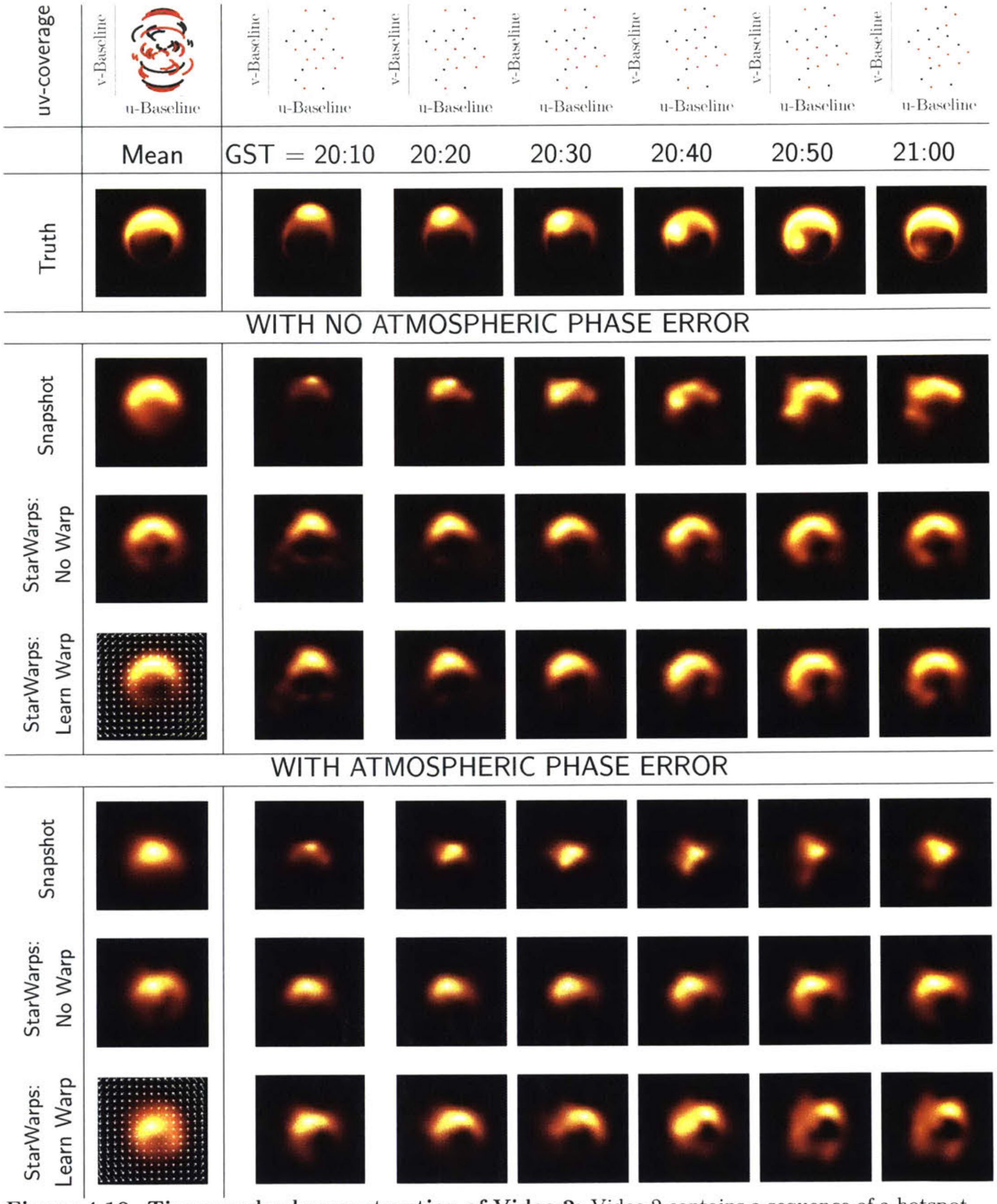


Figure 4.10. Time-resolved reconstruction of Video 2: Video 2 contains a sequence of a hotspot orbiting counter-clockwise around a black hole. We present time-resolved results obtained using data derived from this sequence. Below the true images, we show a subset of images from the baseline 'snapshot imaging' method and compare it to our StarWarps reconstructed video obtained assuming a static warp field or an inferred warp field. The mean image for each sequence is shown in the leftmost column. If we simultaneously estimate a warp field, we indicate the resulting field as arrows on the mean image. Our method substantially improve results over the snapshot method, especially in the case of atmospheric error when the absolute position of the source cannot be recovered. Additionally, despite the fact that this hotspot video does not match our assumed motion model, using our proposed approach we were able to estimate a warp field that provides the direction of the source's true underlying motion. See the caption of Figure 4.9 for more detail.

reconstructions is higher. Furthermore, in the case of data containing atmospheric phase errors our method shows substantial improvement over snapshot imaging. As the closure phase and bispectrum are invariant to the absolute position of the source, each snapshot reconstruction produces an image that is shifted by a different amount. This makes it challenging to align the snapshot frames to pull out meaningful structure in the reconstructed video when there is sparse uv-coverage. For this reason our method substantially outperforms snapshot imaging.

■ 4.5 Conclusion

Traditional interferometric imaging methods are designed under the assumption that the target source is static over the course of an observation [124]. However, as we continue to push instruments to recover finer angular resolution, this assumption may no longer be valid. For instance, the innermost orbital periods around the Milky Way's supermassive black hole, Sgr A*, are just minutes [72]. In these cases, we have demonstrated that traditional imaging methods often break down.

In this work, we propose a way to model VLBI measurements that allows us to recover both the appearance and dynamics of a rapidly evolving source. Our proposed approach, StarWarps, reconstructs a video rather than a static image. By propagating information across time, it produces significant improvements over conventional approaches to create static images or a series of snapshot images in time.

StarWarps uses a multivariate Gaussian imaging regularizer for interferometric imaging, which enables a straightforward optimization method that propagates information through time. In [71], we simultaneously develop alternative methods for reconstructing video from interferometric data. These methods allow for greater flexibility to incorporate a variety of imaging assumptions. However, they are prone to local minima, and thus come at the expense of much more difficulty in converging to the true underlying structure. These approaches have significantly different strengths and may ultimately lead to hybrid approaches for video reconstruction that produce higher quality results, even with noisy and sparse data.

Our technique will hopefully soon allow for video reconstruction of sources that change on timescales of minutes, allowing a real-time view of the most energetic and explosive events in the universe.

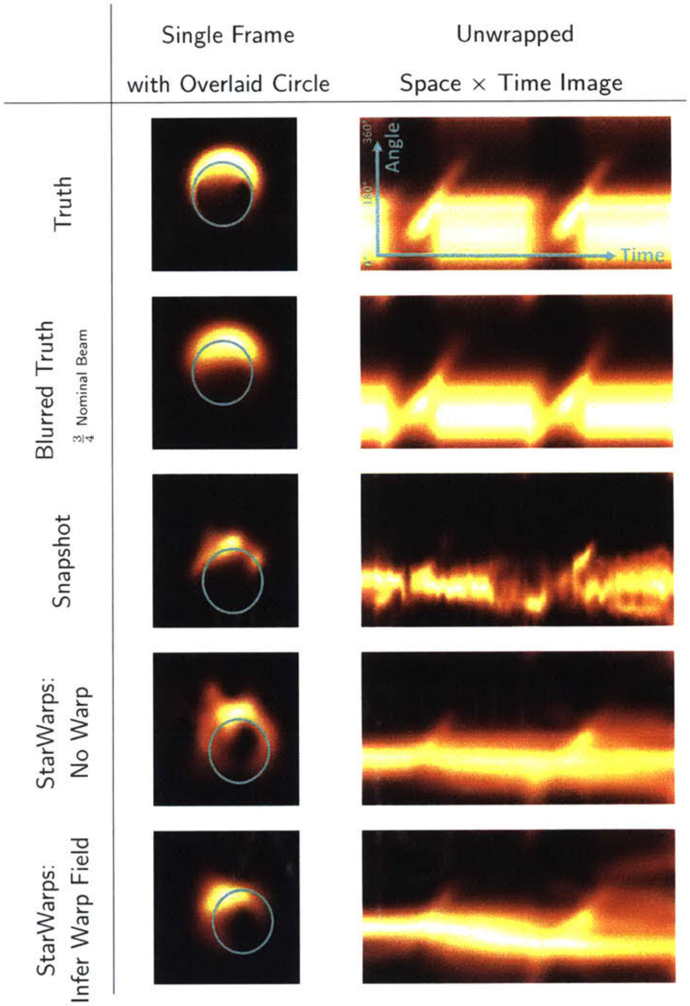


Figure 4.11. Visualizing Recovered Motion: We visualize the recovered motion in Video 2 by displaying the change in intensity around a circle in the image over time. After fitting a circle of constant radius to each video, the intensities around the circle in each image are unwrapped and placed in a single column in the unwrapped space × time image. As the hot spot rotates around the black hole a distinctive line appears in the true angle × time image. These lines also appear in the StarWarps angle × time images, but are harder to discern among the other artifacts in the snapshot imaging result. Results were obtained using the EHT2017+ array with added atmospheric noise, and correspond to results shown in Figure 4.10. As the absolute position of the source is lost when using the closure phase or bispectrum, the position of the recovered black hole moves slightly over the course of the video. This causes the fluctuation in the intensity of the bright horizontal line in the StarWarps recovered angle × time images, as we do not shift the position of the fitted circle.

5

Reconstruction Quality Evaluation & the EHT Imaging Challenges

THERE are many ways to design a VLBI imaging algorithm, and each choice made affects the final reconstruction [17]. For instance, the image prior or regularizer imposed can significantly impact the final appearance of a reconstruction. However, there are many additional choices that can also have an equal impact: the data constraints used, the noise model assumed, the optimization technique employed, and even, as we saw in Chapter 4, the time-evolution assumptions you make about the underlying image.

Within the computer vision and machine learning communities, datasets have become an integral part of research. The introduction of large-scale datasets in fields such as object detection have been the chief reason for the field's significant progress over the last decade [126]. Not only do these datasets provide a large amount of training data, but they also facilitate comparisons between different methods.

Current interferometry testing datasets are small and have noise properties unsuitable for radio wavelengths [10, 13, 83]. In response, in Section 5.1, we introduce a large, realistic VLBI dataset website to the community (`vlbiimaging.csail.mit.edu`). This website allows researchers to easily access a large VLBI dataset, and compare their algorithms to other leading methods.

Furthermore, in preparation for data from the EHT we have developed ongoing challenges to test imaging algorithms. These challenges, referred to as the *EHT Imaging Challenges*, aim to realistically simulate the full imaging and evaluation pipeline before having to image and analyze real EHT data. By running our developed imaging methods through this pipeline, we have been able to better understand each of the algorithms' strengths and weaknesses, and even develop stronger methods. We discuss the EHT Imaging Challenges and the results of the first four challenges in Section 5.2.

5.1 VLBI Reconstruction Dataset

We introduce a dataset and website (`vlbiimaging.csail.mit.edu`) for evaluating the performance of VLBI image reconstruction algorithms. By supplying a large set of easy-to-understand training and testing data, we hope to make the problem more accessible to those less familiar with the VLBI field. The website provides a:

- Standardized data set of real and synthetic data for training and blind testing of VLBI imaging algorithms
- Automatic quantitative evaluation of algorithm performance on realistic synthetic test data
- Qualitative comparison of algorithm performance
- Online form to easily simulate realistic data using user-specified image and telescope parameters

5.1.1 Synthetic Measurements

We provide a standardized format [102] dataset of over 5000 synthetic VLBI measurements corresponding to a variety of array configuration, source images, and noise levels. Measurements are simulated using the MIT Array Performance Simulator (MAPS) software package (see Section 3.4.1) [85].

We generate data using a collection of black hole [23], astronomical [1, 2], and natural images. We have deliberately included a diversity of images in the imaging database, since imaging algorithms for black holes must be sufficiently non-committal that they can identify departures from canonical expectations. Natural images test robustness to complex scenes with varied image statistics.

Generate Your Own Data

Another useful feature of this website is the ability to generate your own synthetic VLBI data easily from an image. This feature makes it easy to test algorithms under arbitrary configurations, and can be very valuable when a researcher is working on imaging with a telescope array not used in the training data described in Section 5.1.1. An online form allows anyone to generate realistic data with a preselected or uploaded image. This form requires the user to specify the location of the source, the telescope array, the time of the observation, data processing information, and the types of noise introduced into the measurements. Generated data using these properties is then made readily available for download.

5.1.2 Real Measurements

We provide 33 sets of measurements from the VLBA-BU-BLAZAR Program [73] in the same standardized format [102]. This program has been collecting data on a number of blazars every month since 2007. Measurements are taken using the Very Long Baseline Array (VLBA) at 43 GHz. Although both the angular resolution and wavelength of these measurements are very different from those taken by the EHT (which collects at $\approx$ 230 GHz) [44], they provide a means to test algorithms on measured, experimental data.

5.1.3 Automatic Evaluation System

As part of the dataset website we also provide an automatic evaluation system that facilitates unbiased comparisons between algorithms, which are otherwise difficult to make and are lacking in the literature. In particular, we introduce a blind test set of challenging synthetic data. Measurements with realistic errors are generated using a variety of target sources and telescope parameters and provided in the UVFITS and OIFITS format [102]. Researchers are encouraged to run their algorithms on this data and submit results to the website for evaluation.

Traditional point-by-point error metrics, such as MSE and PSNR, are sometimes uninformative in the context of highly degraded VLBI reconstructions. Therefore, we supplement the MSE metric with the perceptually motivated structural similarity (SSIM) index [131]. Since the absolute position of the emission is lost when using the bispectrum, we first align the reconstruction to the ground truth image using cross-correlation. We then evaluate the MSE and SSIM on the normalized, aligned images. Although we consider MSE and SSIM a good first step towards quantitative analysis, we believe a better metric of evaluation is subject for future research.

5.2 Event Horizon Telescope Imaging Challenges

EHT observations are both extremely sparse and difficult to calibrate, thus traditional radio-interferometric imaging methods often perform poorly. Consequently, in preparation for spring 2017 observations with the EHT, there has been a significant effort to develop imaging methods that are capable of robustly reconstructing images under these challenging conditions. Although there has been considerable progress made in VLBI imaging during the last couple of years, there are still a number of questions that remain largely unanswered.

- **What are the pros and cons of different imaging methods?** We know each algorithm has its own biases and is good at reconstructing different structures, but sometimes its hard to figure out exactly what those biases are when each method is evaluated on different datasets.
- **What makes a successful imaging algorithm?** Is it the prior that is the most important component of an algorithm, or the way that data constrains the reconstruction process? Answering these questions will help us improve our current imaging methods.
- **What should we expect with real EHT data?** Even if the authors of imaging algorithms try very hard to perform fair comparisons, there is always sort of bias when they evaluate their own algorithms on data that they personally generated. If we mitigate these biases do algorithms still perform at the same level, and what conclusions can we draw if we dont get identical resulting images from each algorithm?

To help answer these questions, and to establish confidence in our forthcoming reconstructions from EHT data, it is imperative that we rigorously evaluate the methods under a common set of realistic circumstances. The Event Horizon Telescope Imaging Challenges provide a way to do this while also preemptively testing out the EHT's imaging pipeline. These challenges not only provide an opportunity for us to better understand each of the imaging algorithms' strengths and weaknesses, but also provide guidance in developing stronger methods.

In this section we summarize results from the first four Event Horizon Telescope Imaging Challenges. These challenges consisted of blind imaging of different synthetic data sets corresponding to realistic measurements from the EHT. Challenge participants were provided with details of telescope and noise properties, but the true emission images/videos used to generate the data remained undisclosed. These challenges have involved a diverse set of researchers with expertise in a variety of specialties. In addition to those who work on imaging algorithms, researchers who generate images, understand how to model realistic noise in the data, and can provide expertise when it comes to an interpretation of the resulting images were all essential to the success of these challenges.

In Section 5.2.1 we discuss the general pipeline of the imaging challenges. In Section 5.2.2 we discuss data generation and what types of noise/error was included in the synthetic measurements. In Section 5.2.3 we describe the forms of quantitative and qualitative analysis done for each challenge. In Section 5.2.4 we give details of each imaging challenge, display the submissions, and provide analysis of the results. In Section 5.2.5 we provide concluding remarks, including a discussion on lessons we have learned thus far from these challenges.

■ 5.2.1 Challenge Pipeline

The challenge pipeline consists of a set of circular steps (See Figure 5.1):

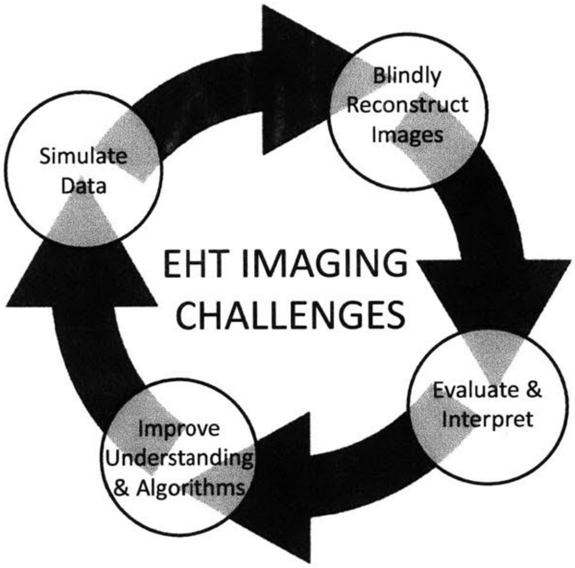


Figure 5.1. The EHT Imaging Challenge Pipeline: The imaging challenges consist of four steps: simulation, reconstruction, evaluation, and learning. This process encourages algorithm development and improves our understanding of imaging methods on EHT quality data. This newly acquired knowledge then helps to shape the next imaging challenge.

1. **Simulate Data:** A team of researchers simulate realistic datasets corresponding to measurements made by the EHT. These datasets are publicly distributed on `http://vlbiimaging.csail.mit.edu/imagingchallenge` without information about the truth image or video used to generate data.

2. **Blindly Reconstruct Images:** Teams or individuals independently reconstruct images from the publicly available data, and submit their reconstructions to the website.

3. **Evaluate & Interpret:** The submitted images are sent to a "panel of experts" who try to predict what the underlying emission is caused by, and qualitatively evaluates the images without knowledge of the ground truth image/video. Submitted images are also quantitatively compared to the true image/video using standard fidelity metrics.

4. **Improve Understanding & Algorithms:** The performance of imaging methods and the evaluation process are assessed. This assessment helps us to improve the imaging algorithms and pinpoint issues with the current imaging pipeline.

■ 5.2.2 Synthetic Data Generation

To produce each dataset, visibilities were synthetically generated that corresponded to realistic EHT observations of known brightness distributions on the sky (truth images) at a wavelength of 1.3 mm. Generating synthetic data consists of three steps: selecting the true image or video, generating visibilities corresponding to telescopes in the EHT array, and adding realistic noise to the visibilities.

Data was provided in a variety of formats (e.g., UVFITS, OIFITS, and text files). Information about the location of each emission region, the telescopes used, and the types of noise added were always provided to the participants. Information about the total flux of the source was provided in the case of static emissions, and the total field of view was only provided for a few of the datasets.

Selecting the True Image or Video

The first step in generating a synthetic VLBI dataset is choosing the underlying source image that will be observed. We want to make sure the imaging algorithms can reliably reconstruct an image of the black hole, even if it does not match our expectations. For this reason, many different kinds of images have been used in the challenges thus far. In addition to a varied set of black hole simulation images, we also have used real images of other astronomical sources (e.g. blazars), and a non-astronomical terrestrial image.

Both static and time-varying sources were tested in the imaging challenges. In the first four challenges, there were eight datasets containing static sources and two datasets of time-varying sources. Specifics of each truth image or video is discussed in the subsequent sections of this chapter.

Generating EHT Visibilities

To generate realistic visibilities, each truth image's Fourier transform was computed along projected baselines corresponding to telescopes in the EHT. Specifically, a subset of the telescopes located in Hawaii, Arizona, Mexico, Chile, France, Spain, and the South Pole were used. The coordinates and SEFD that were assumed for each telescope are specified in Table 2.2 in Chapter 2.

In the case of a static emission, the same image was used to compute each set of measurements for the full observation. In the case of a time-varying emission, a single frame of a video was used to obtain the set of measurements corresponding to a single time during the observation.

Adding Noise to the Visibilities

Three different types of noise were added to the ideal complex visibilities: thermal, atmospheric (randomized station phases), and systematic (gain calibration errors). Furthermore, as time-varying emissions break common assumptions used in VLBI imaging, we also categorize time-variability as a form of "noise". By providing challenges with different kinds of noise, we are able to study how robust different methods are to varying levels and types of noise.

Thermal Noise: Isotropic Gaussian noise was added to the simulated complex visibilities, Γ, to mimic the effect of thermal noise. Following the characterization of thermal noise presented in Equation 2.24 of Chapter 2, the standard deviation of the Gaussian noise for a visibility was fixed based on bandwidth ($\Delta\nu$), integration time (τ), and each telescope's System Equivalent Flux Density(SEFD). A factor of $\eta = 1/0.88$ was included due to the effect of 2-bit quantization [124].

Atmospheric Noise: As discussed in Section 2.3.1 of Chapter 2, variable amounts of water vapor in the atmosphere above each telescope site cause there to be large fluctuations in the propagation delay of light at 1.3 mm. The atmosphere affects an ideal visibility, Γ, by introducing an additional, rapidly varying, station-based phase term:

$$\Gamma_{j,k}^{\text{measured}} = \exp\left[i(\phi_j(t) - \phi_k(t))\right] \, \Gamma_{j,k}^{\text{ideal}} \tag{5.1}$$

where ϕ_k and ϕ_j are the phase delays introduced in the path to telescopes k and j respectively. To simulate atmospheric errors, a phase shift was introduced in each complex visibility [124]. This phase shift incorporated random phases drawn uniformly from $[0, 2\pi)$ for each corresponding site at each time step, t.

Systematic Gain Miscalibration Error: Systematic (site-based) gain errors are introduced as error in the SEFD used for calibration. Gain calibration errors are modeled at each site as a time-dependent gain $G(t)$ and a time-dependent opacity $p(t)$. Given the source elevation at time t, $\theta(t)$, a time-dependent SEFD is computed using the relationship

$$\mathrm{SEFD}(t) = \mathrm{SEFD}^{\mathrm{ideal}} \; \frac{\exp\left[p(t)\right] / \sin\left[\theta(t)\right]}{G(t)} \tag{5.2}$$

This new time-dependent SEFD measure is used in place of the measured SEFD when computing thermal noise (Refer to Equation 2.24) on each visibility [26]. The gain calibration and opacity are both modeled as Gaussian random processes with 1-sigma values equal to:

- 10% constant SEFD variation for each site, plus 10% additional scan-by-scan variation
- 10% variation in the true opacity

When systematic gain errors are introduced into a dataset, the visibility amplitudes often exceed the emission's total flux density (see Figures 5.6, 5.18, and 5.24).

5.2.3 Analysis and Evaluation

We evaluate the accuracy of each of the reconstructions using a variety of methods. First, we perform a traditional quantitative analysis using point-by-point fidelity metrics. Second, we provide feedback on the reconstructions from a panel of experts who were asked to evaluate the reconstructed images without knowledge of the truth images.

Quantitative Analysis

We evaluate the performance of each submitted reconstruction using traditional point-by-point fidelity metrics: normalized root-mean-square-error (normalized RMSE) and normalized cross-correlation.

Blur Levels To analyze the agreement of submitted reconstructions with the truth image at a variety of resolutions, each of the images was blurred with a series of anisotropic restoring beams. The nominal beam size was computed for each challenge by fitting a Gaussian to the dirty beam [26]. Once the nominal beam size was determined, each of the truth and submitted reconstruction images were convolved with a series of beams that were a fraction of the nominal beam size. Specifically, each image was convolved with a beam of $k/10$ times the nominal beam size for $k \in \{0, ..., 15\}$. We refer to the convolved truth and reconstruction images as *c-truth* and*c-reconstruction* images, respectively.

Figures 5.5, 5.8, 5.11, 5.14, 5.17, 5.20, 5.23, 5.26, 5.29, and 5.34 show a sample of c-truth images for each of the ten datasets. Note that the beam size in pixels depends not only on the telescope array, but also the image's field of view.

Registration To evaluate the performance of a c-reconstruction using point-by-point error metrics, the c-reconstruction must be properly aligned with the c-truth image. However, methods that rely on closure quantities (e.g. bispectrum, closure phases) are often invariant to absolute position. Therefore, these reconstructions are generally translated from the true position.

To correct for this translation before evaluating the reconstruction error, each c-reconstruction was registered to its corresponding c-truth image. This is done by maximizing the cross-correlation between the normalized images. Alignments were done on images padded to have twice the FOV with a pixel spacing specified by the truth image.

Fidelity Metrics We evaluate the normalized root-mean-squared-error (RMSE) and normalized cross-correlation for each independently registered c-reconstruction. These metrics are defined below for an N pixel c-reconstruction, r_c, and its corresponding c-truth, g_c, image.

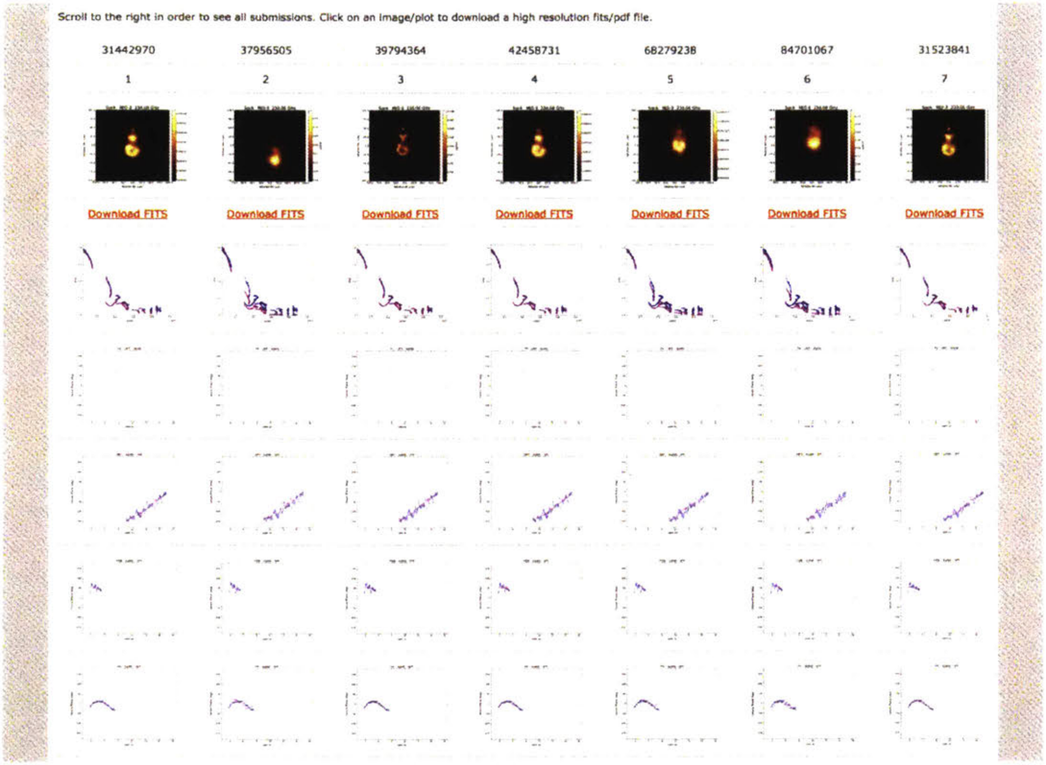
$$\text{Normalized RMSE}_c = \sqrt{\frac{\sum_n |r_c[n] - g_c[n]|^2}{\sum_n |r_c[n]|^2}} \tag{5.3}$$

$$\text{Normalized Cross-Correlation}_c = \frac{1}{N} \sum_n \frac{(r_c[n] - \bar{r}_c) \times (g_c[n] - \bar{g}_c)}{\sigma_{r_c} \sigma_{g_c}} \tag{5.4}$$

The terms $\bar{x}$ and σ_x indicate the mean and standard deviation of values in x respectively.

Qualitative Analysis

An independent panel of two to three experts were also asked to review the reconstructions – without knowledge of the truth images – and provide feedback. The goal of this exercise was to get a quick sense of how we might actually be able to use the reconstructed images to infer the structure of the underlying source, and to assess the reliability of these conclusions.



Scroll to the right in order to see all submissions. Click on an image/plot to download a high resolution fits/pdf file.

Figure 5.2. A screen shot of the evaluation website used by judges in the fourth challenge (dataset 7). Judges were provided with the submission images, challenge data, and plots that showed how well each image fit the simulated measurements. Additional plots (such as the uv-coverage, beam size, etc) were provided upon request.

Each expert judge was provided with a website containing the submissions for each dataset, along with plots showing how well the submitted image aligned with the measured amplitude and closure phases. Refer to Figure 5.2 for a sample screen shot. These plots made it easier for the panel to compare the reconstructed images and to analyze how well each image fit the provided data in amplitude and closure phase. Judges were welcome to ask for additional materials to help them in the judging process, but were not given any information about the underlying truth image.

Nine different judges participated on the panels: Sheperd Doeleman, Vincent Fish, Laurent Loinard, Mark Reid, Motoki Kino, Fabien Baron, Jim Moran, Charles Gammie, and Heino Falcke. In each dataset section we provide a few comments from each associated judge.

5.2.4 Challenge Datasets & Results

The imaging challenges thus far have consisted of 4 different challenges containing 10 different datasets. In Table 5.1 we summarize the conditions of each dataset, and in the following sections we discuss each dataset in further detail along with results.

Challenge - Dataset	Location	Telescopes	Flux (Jy)	FOV (μas)	Noise Property
1 - 1	Sgr A*	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, PDB, SPT	2	340	Thermal
1 - 2	M87	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, PDB, SPT	2	140	Thermal, Atmospheric, Systematic
1 - 3	Sgr A*	SMA, SMT, LMT, ALMA, SPT	2	500	Thermal, Atmospheric
2 - 4	3C279	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, PDB, SPT	3	-	Thermal, Atmospheric
2 - 5	M87	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, PDB, SPT	2	-	Thermal, Atmospheric
2 - 6	Sgr A*	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, PDB, SPT	2	-	Thermal, Atmospheric Systematic
3 - 7	Sgr A*	SMA, SMT, LMT, ALMA, PV, PDB, SPT	2	200	Thermal, Atmospheric
4 - 8	Sgr A*	SMA, SMT, LMT, ALMA, PV, PDB, SPT	2	-	Thermal, Atmospheric Systematic
4 - 9	Sgr A*	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, SPT	-	-	Thermal, Atmospheric Time-Variability
4 - 10	Sgr A*	SMA, JCMT, SMT, LMT, ALMA, APEX, PV, SPT	-	-	Thermal, Atmospheric Time-Variability

Table 5.1. Challenge Conditions: These specified conditions used to generate challenge datasets. All information in this table was provided to challenge participants. Specifics of each telescope can be seen in Table 2.2 of Chapter 2.

Challenge 1 - Dataset 1

Image: The image used to generate measurements for this dataset was a ray-traced result from a GRMHD simulation contributed by Chi-Kwan Chan [27]. The system exhibits a faint jet and prominent black hole shadow. The total image field of view is 340 μ-arcseconds and the total flux density is 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), France (PDB), Spain (PV), and the South Pole (SPT). The telescopes were assumed to be pointed towards Sgr A*and only thermal noise was included in the visibility measurements. Thus, there were no phase corruptions due to the atmosphere. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.3. In this dataset both the field of view and the total flux of the source were provided to the participants.

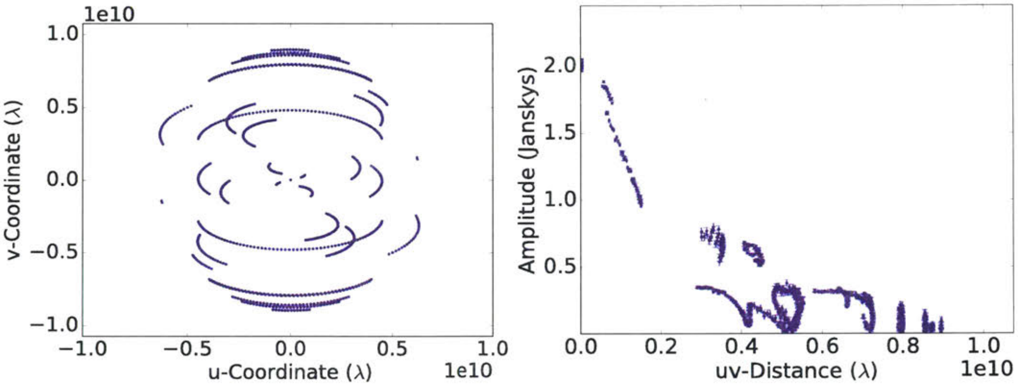
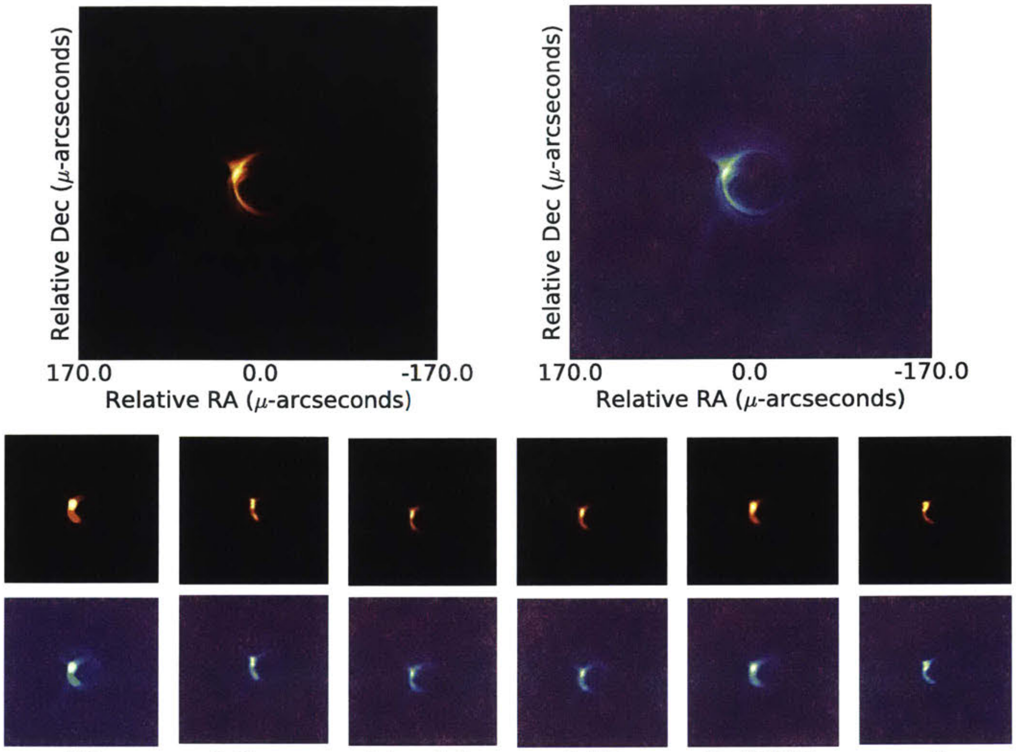


Figure 5.3. Dataset 1: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.

Submissions: Six submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.4 for the two colormaps along with the truth image. The method names and closest appropriate references for each submission are listed below:

- Akiyama & Tazaki using Sparse Modeling with Complex Visibilities (A&T: Visibility SM) [62]
- Chael using MEMHorizon with Closure Amplitudes and Closure Phases (C: Closure Only MEMHor) [26]
- Chael using MEMHorizon with Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]



A&T: Visibility SM

C: Closure Only MEMHor

C: Amp+Cl-Ph MEMHor

C: Bisp MEMHor

Bo: Visibility CHIRP

D: PyMORESANE

Figure 5.4. Dataset 1: Truth (top) and submitted images in two colormaps.

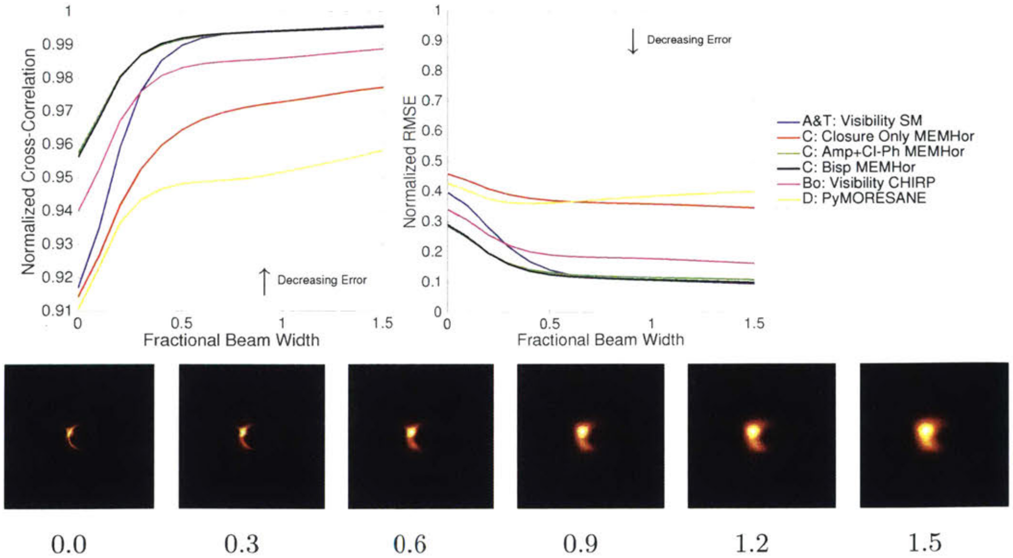


Figure 5.5. Dataset 1: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Chael using MEMHorizon with the Bispectrum (C: Bisp MEMHor) [26]
- Bouman using CHIRP with Complex Visibilities (Bo: Visibility CHIRP)
- Deane using PyMORESANE (D:PyMORESANE) [30]

Evaluation A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.5. Although all submitted images were very consistent with the data, under these measures C: Amp+Cl-Ph MEMHor and C: Bisp MEMHor performed the best.

Qualitative analysis was done by a panel of three judges. All judges expressed results were very consistent, and believed there was strong evidence that the underlying image was a black hole shadow. All panelists selected C: Amp+Cl-Ph MEMHor as a top reconstruction choice. We provide selected quotes from each panelist below:

> This clearly looks like a black hole shadow with a crescent shape due to Doppler boosting on one side. It appears that the rotation/orbital axis is roughly NS with left side approaching. The diameter of the shadow would be about 50uas, but it's hard to read off the figures - I imagine that fitting to the image would give a pretty robust estimate. -Sheperd Doeleman

> The reconstructions are mostly consistent with each other. If we reconstructed this set of images, we'd be pretty confident publishing one or more of these images. -Vincent Fish

> If there is a black hole, the diameter of the shadow would be about 50 uas. -Laurent Loinard

Challenge 1 - Dataset 2

Image: The image used to generate measurements was a ray-traced result from a GRMHD simulation contributed by Roman Gold [53]. This image shows a faint black hole shadow feature but with a more prominent circular feature from emission along the jet funnel wall. The total image field of view was 140 μ-arcseconds and the image's total flux density was rescaled to 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), France (PDB), Spain (PV), and the South Pole (SPT). The telescopes were assumed to be pointed towards M87. Thermal noise, atmospheric phase error, and gain mis-calibration error were all included in the visibility measurements. The uv-coverage and the visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.6. In this dataset both the field of view and the total flux of the source were provided to the participants.

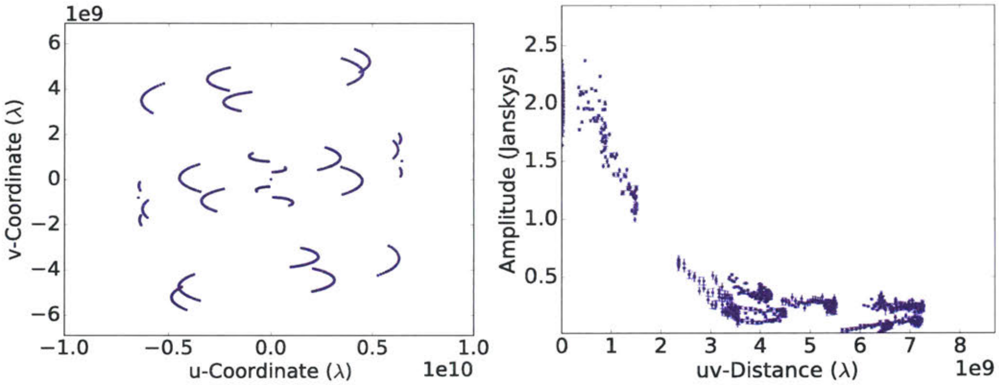
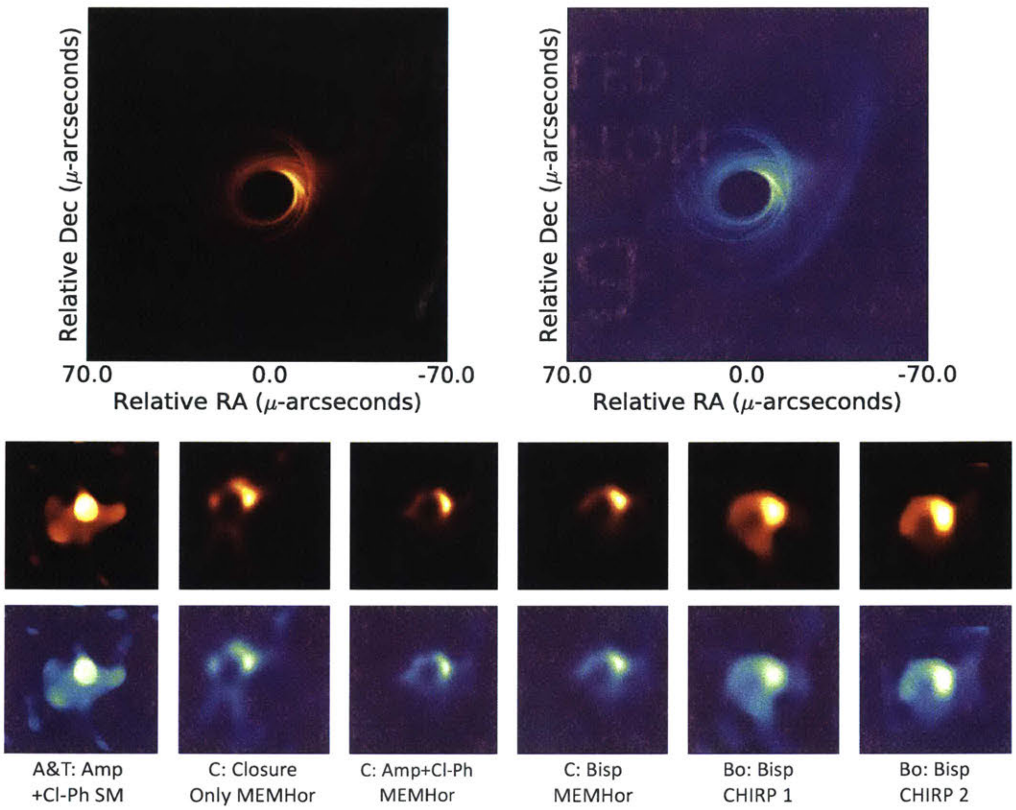


Figure 5.6. Dataset 2: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that amplitude mis-calibration errors cause the amplitude to sometimes exceed the emission's total flux of 2 Jy.

Submissions: Six submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.7 for the two colormaps along with the truth image. The method names and closest appropriate references for each submission are listed below:

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM) [6, 62]
- Chael using MEMHorizon with Closure Amplitudes and Closure Phases (C: Closure Only MEMHor) [26]



A&T: Amp +Cl-Ph SM

C: Closure Only MEMHor

C: Amp+Cl-Ph MEMHor

C: Bisp MEMHor

Bo: Bisp CHIRP 1

Bo: Bisp CHIRP 2

Figure 5.7. Dataset 2: Truth (top) and submitted images in two colormaps.

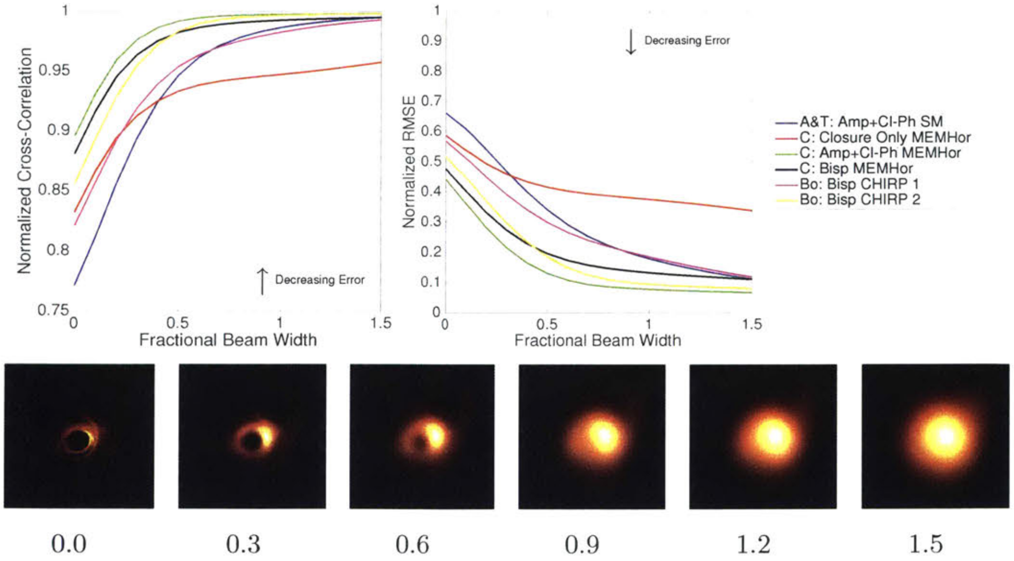


Figure 5.8. Dataset 2: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Chael using MEMHorizon with Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Chael using MEMHorizon with the Bispectrum (C: Bisp MEMHor) [26]
- Bouman using CHIRP with the Bispectrum (Bo: Bisp CHIRP 1 & 2)

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.8. Under both metrics C:Amp+Cl-Ph MEMHor fit the true image the best.

Qualitative analysis was done by a panel of three judges. Although the panelists identified a ring structure in many of the reconstructions, there was concern about the consistency of the results. All three panelists chose C: Amp+Cl-Ph MEMHor as a top reconstruction. We provide selected quotes from each panelist below.

> ... appears to be a 'ring' of some sort, but smaller (20uas diameter)
> -Sheperd Doeleman

> This dataset may be badly miscalibrated. The reconstructions are inconsistent with each other. If we reconstructed this set of images from the same set of data, I would vote against publishing any of them. -Vincent Fish

> If there is a black hole, the diameter of the shadow would be about 30 uas. -Laurent Loinard

Challenge 1 - Dataset 3

Image: This dataset's image was created by adopting a reconstructed image of 3C273 from 7 mm VLBI by the BU Blazar group [73]. The particular epoch chosen was 4 July, 2012. This image shows a prominent one-sided jet extended from a bright core, as is characteristic in many longer-wavelength VLBI images. The original image had a field of view of approximately 5 mas, so the image was rescaled by a factor of 1/10 to give a resulting field of view of approximately 500 μas. The image was also rotated from its original orientation. The total flux density of the image was rescaled to 2 Jy.

Data: This dataset was generated using a smaller telescope array with only sites in Hawaii (SMA), Arizona (SMT), Mexico (LMT), Chile (ALMA), and the South Pole (SPT). Note that JCMT and APEX were not included in the array, so there were no short baselines to help with calibration. The telescopes were assumed to be pointed towards Sgr A*and and both thermal noise and atmospheric phase error was included in the visibility measurements. The uv-coverage and the visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.9. In this dataset both the field of view and the total flux of the source were provided to the participants.

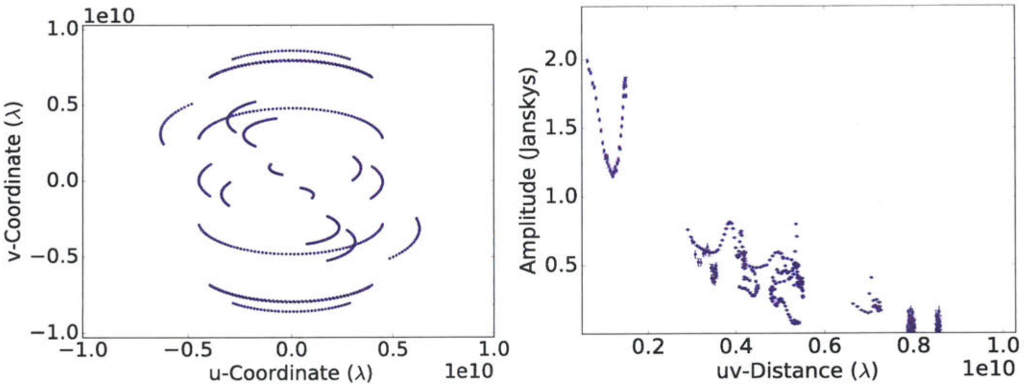
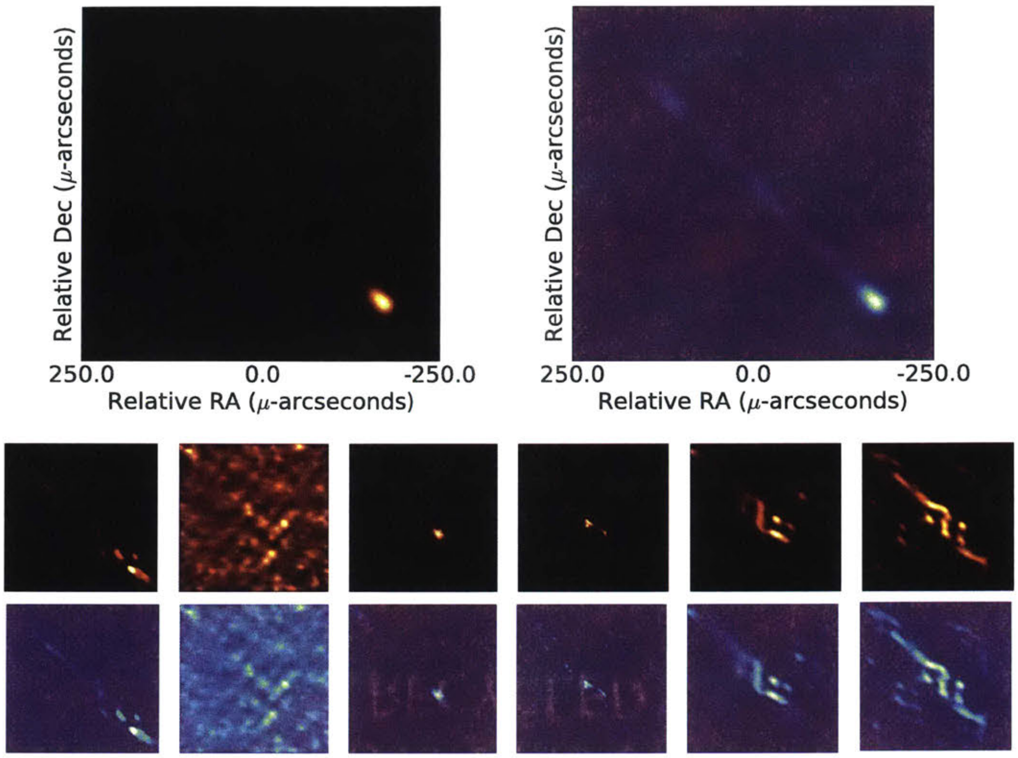


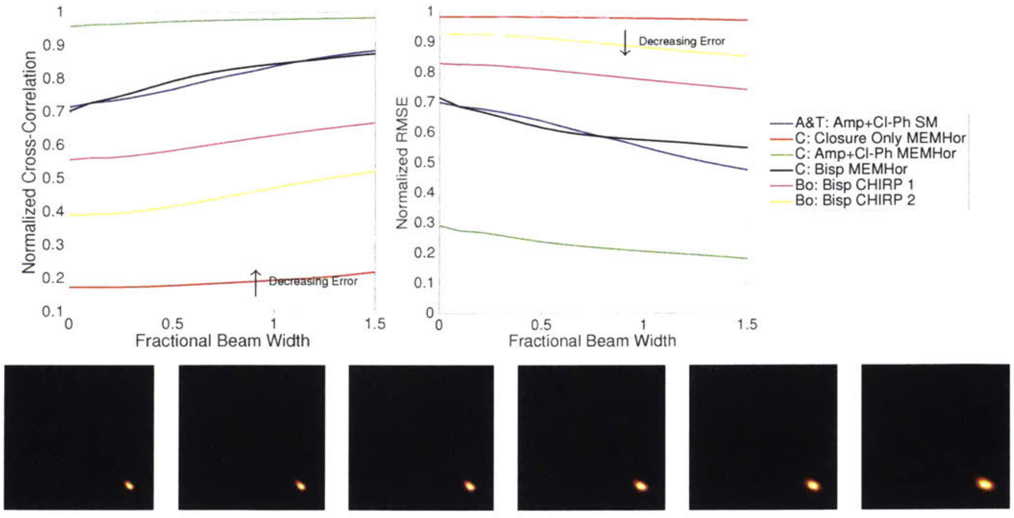
Figure 5.9. Dataset 3: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.

Submissions: Six submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.10 for the two colormaps along with the truth image. The method names and closest appropriate references for each submission are listed below:



A&T: Amp +Cl-Ph SM | C: Closure Only MEMHor | C: Amp+Cl-Ph MEMHor | C:Bisp MEMHor | Bo: Bisp CHIRP 1 | Bo: Bisp CHIRP 2

Figure 5.10. Dataset 3: Truth (top) and submitted images in two colormaps.



0.0 | 0.3 | 0.6 | 0.9 | 1.2 | 1.5

Figure 5.11. Dataset 3: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM) [62]
- Chael using MEMHorizon with Closure Amplitudes and Closure Phases (C: Closure Only MEMHor) [26]
- Chael using MEMHorizon with Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Chael using MEMHorizon with the Bispectrum (C: Bisp MEMHor) [26]
- Bouman using CHIRP with the Bispectrum (Bo: Bisp CHIRP 1 and Bo:Bisp CHIRP 2)

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.11. Under both metrics C:Amp+Cl-Ph MEMHor fit true image the best by a large margin.

Qualitative analysis was done by a panel of three judges. Although the panelists all felt the results were inconsistent, they correctly identified the underlying image as being a jet. All three panelists chose C: Amp+Cl-Ph MEMHor as a top reconstruction. We provide selected quotes from each panelist below.

> Hard to tell what this might be. Some discrete points or a filamentary structure - jet?. The big dip in amplitude on the short baseline means there is some extended structure on 1G-lambda scales (200 uas). I guess I'd say this is a core-jet with the jet going off to the NE (upper left). -Sheperd Doeleman

> At least three of the reconstructions converge on something like a one-sided jet with a bright core. -Vincent Fish

> I dont see strong evidence for a shadow. Its interesting that there is a wide variety of reconstructed images. From non-detection to a dominant compact central object, with perhaps a jet, to a complex funny looking structure... -Laurent Loinard

Challenge 2 - Dataset 4

Image: This dataset's image was created by adopting a reconstructed image of 3C279 from 7 mm VLBI by the BU Blazar group [73]. This image features a prominent jet structure. The particular epoch chosen was the 18th of March, 2016. The original image had a field of view of approximately 3 milli-arcseconds, but was rescaled to a field of view of approximately 500 μ-arcseconds. The image was also rotated from its original orientation. The total flux density of the image was rescaled to 3 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), France (PDB), Spain (PV), and the South Pole (SPT). The telescopes were assumed to be pointed towards 3C279 and both thermal noise and atmospheric phase error were included in the visibility measurements. The uv-coverage and the visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.12. In this challenge the total flux of the source was provided to the participants, but the field of view was left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size.

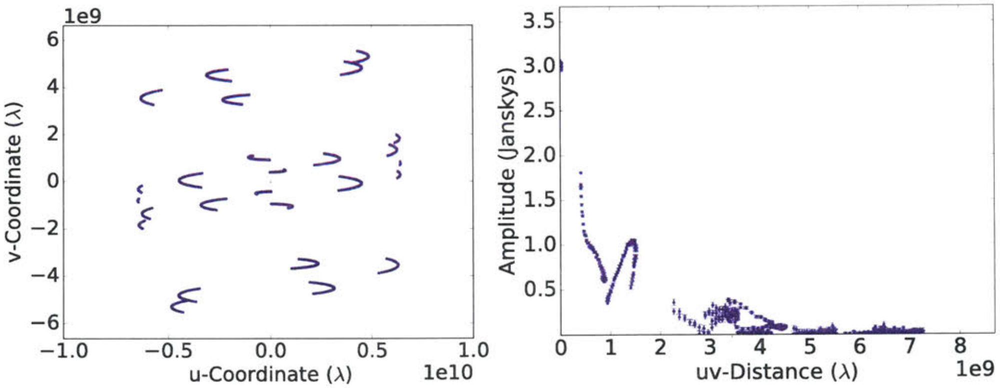
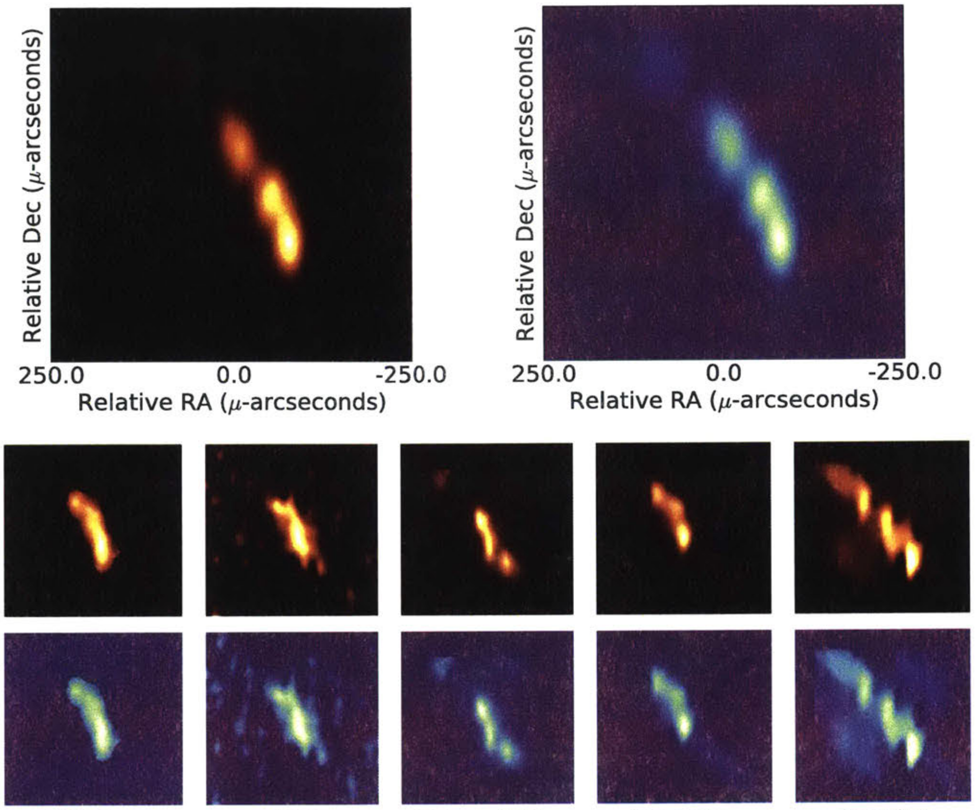


Figure 5.12. Dataset 4: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.

Submissions: Five submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.13 for the two colormaps along with the truth image. The method names and closest appropriate references for each submission are listed below:



A&T: Amp+ Cl-Ph SM

A&T: Self-Cal SM

C: Amp+Cl-Ph MEMHor

Bo: Amp+ Cl-Ph CHIRP

Bo: Bisp CHIRP

Figure 5.13. Dataset 4: Truth (top) and submitted images in two colormaps.

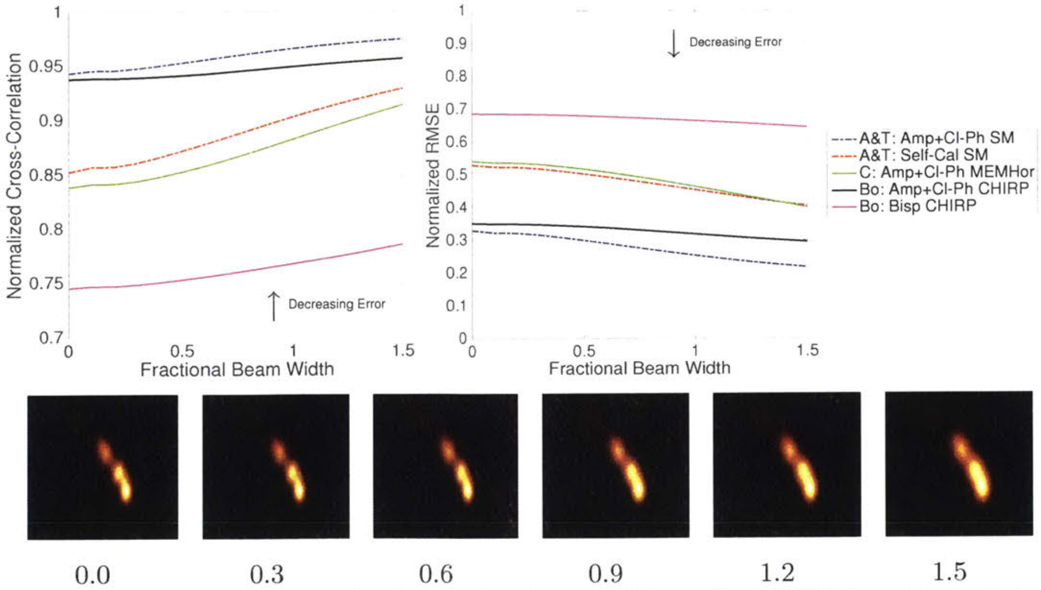


Figure 5.14. Dataset 4: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM) [6, 62]
- Akiyama & Tazaki using Sparse Modeling with Self-Calibrated Visibilities (A&T: Self-Cal SM) [6, 62]
- Chael using MEMHorizon with Visibility Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Bouman using CHIRP with the Visibility Amplitudes and Closure Phases (Bo: Amp+Cl-Ph CHIRP)
- Bouman using CHIRP with the Bispectrum (Bo: Bisp CHIRP)

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.14. Under these metrics, both A&T: Amp+Cl-Ph SM and Bo: Amp+Cl-Ph fit the true image fairly well, although A&T: Amp+Cl-Ph SM performed slightly better, especially when the metrics focused on the broader structure's fit.

Qualitative analysis was done by a panel of three judges. All three panelists chose Bo:Amp+Cl-Ph CHIRP as a top reconstruction. We provide selected quotes from each panelist below.

> If this is a core-jet structure, [assuming Bo: Amp+Cl-Ph CHIRP is correct] there is a hint of a counter jet... [and assuming A&T: Amp+Cl-Ph SM is correct] there is no hint of a counter jet. - Mark Reid

> My guess at the underlying true image: A core-jet going to North with a bend to NE in the middle. -Motoki Kino

> I would say all images appear flaw[ed] in one way or another and would be dismayed if we ever got something like this out of the EHT. -Fabien Baron

Challenge 2 - Dataset 5

Image: The image used to generate measurements for this dataset was a 1.3 mm ray-traced result from a GRMHD simulation of M87 contributed by Monika Mościbrodzka [94]. This jet model was generated assuming a black hole of 3.5 billion solar masses, a lower estimate on the size of M87's black hole. The image is viewed at 30 degrees from the pole, and exhibits a faint jet and prominent black hole shadow. The full image's field of view is 160 μ-arcseconds and the total flux density is scaled to 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), France (PDB), Spain (PV), and the South Pole (SPT). The telescopes were assumed to be pointed towards M87 and both thermal noise and atmospheric phase error were included in the visibility measurements. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.15. In this dataset the total flux of the source was provided to the participants, but the field of view was left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size.

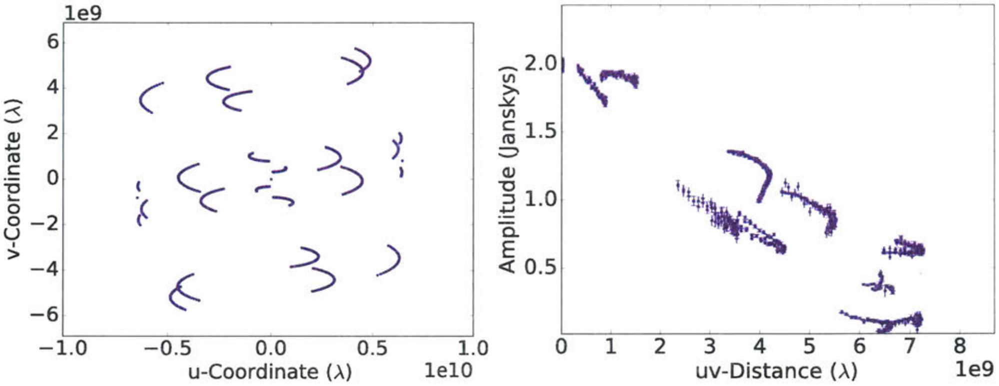
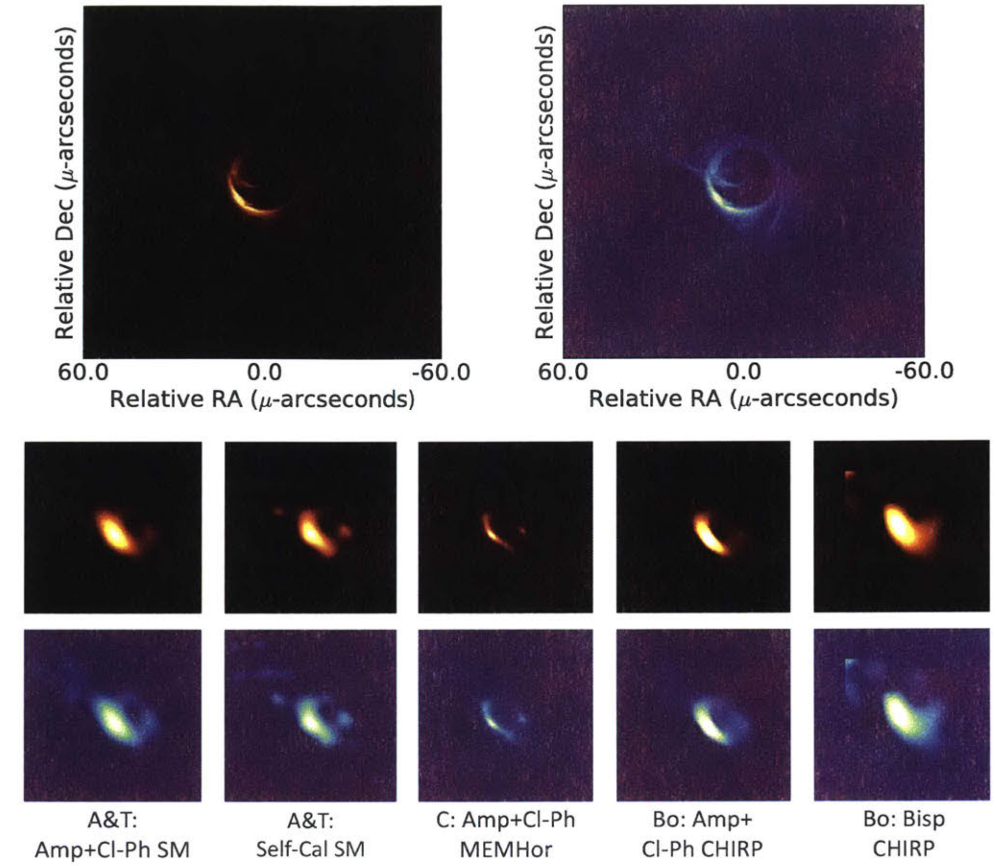


Figure 5.15. Dataset 5: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.

Submissions: Five submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.16 for the two colormaps along with the truth image. The method names and appropriate references for each submission are listed below:



A&T: Amp+Cl-Ph SM

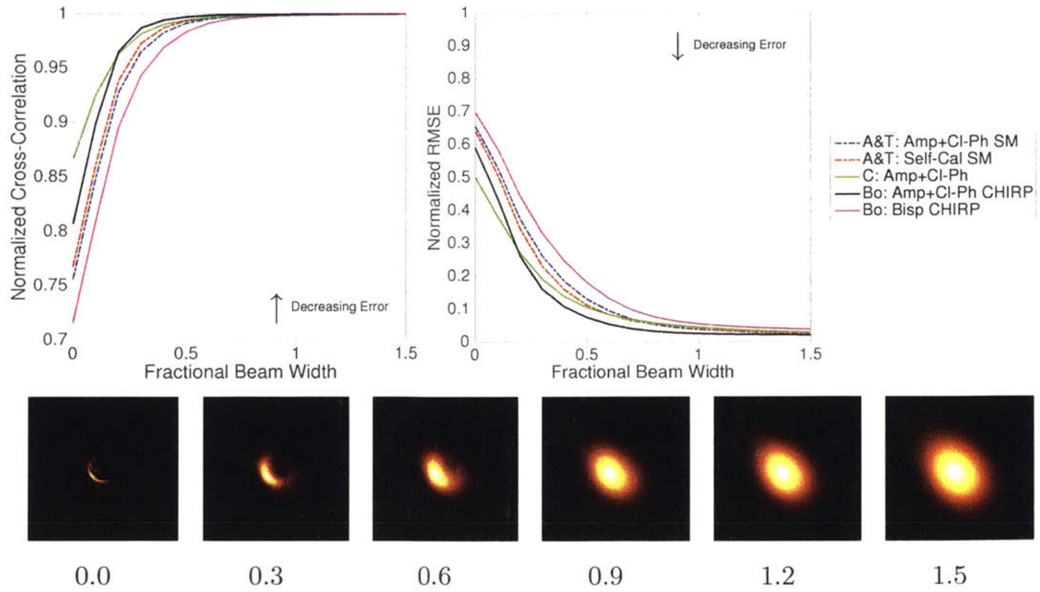
A&T: Self-Cal SM

C: Amp+Cl-Ph MEMHor

Bo: Amp+ Cl-Ph CHIRP

Bo: Bisp CHIRP

Figure 5.16. Dataset 5: Truth (top) and submitted images in two colormaps.



0.0 0.3 0.6 0.9 1.2 1.5

Figure 5.17. Dataset 5: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Tazaki & Akiyama Sparse Modeling using Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM) [6, 62]
- Tazaki & Akiyama Sparse Modeling using Self-Calibration (A&T: Self-Cal SM) [6, 62]
- Chael using Visibility Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Bouman CHIRP using the Visibility Amplitudes and Closure Phases (Bo: Amp+Cl-Ph CHIRP)
- Bouman CHIRP using the Bispectrum (Bo: Bisp CHIRP)

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.17. Under these error metrics all submissions performed well, partly due to the fact that the structure was quite small compared to the beam size. C:Amp+Cl-Ph MEMHor performed the best at super resolving structure of the source. However, after convolving with a beam size of at least 0.2 the nominal beam Bo:Amp+Cl-Ph CHIRP performed slightly better than the rest of the submissions.

Qualitative analysis was done by a panel of three judges. The judges all felt like there was a ring structure, but did not agree on which reconstruction was the best. We provide selected quotes from each panelist below.

> This appears to be a ring-source with enhanced brightness on the left side. - Mark Reid

> My guess at the underlying true image: This is a torus structure. I can say that a hole at the center can be a BH-shadow in astrophysical point of view. -Motoki Kino

> The best-looking images are the ones without bleeding and with a deeper shadow "well"... -Fabien Baron

Challenge 2 - Dataset 6

Image: The image used to generate measurements for this dataset was a 1.3 mm ray-traced result from a GRMHD simulation of Sgr A* contributed by Hotaka Shiokawa [118]. The image of this SANE model is viewed at 45 degrees from the pole, and exhibits a faint ring. The full image's field of view is 160 μ-arcseconds and the total flux density is scaled to 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), France (PDB), Spain (PV), and the South Pole (SPT). The telescopes were assumed to be pointed towards Sgr A*. Thermal noise, atmospheric phase error, and systematic amplitude calibration errors were all included in the visibility measurements. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.18. In this dataset the total flux of the source was provided to the participants, but the field of view was left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size.

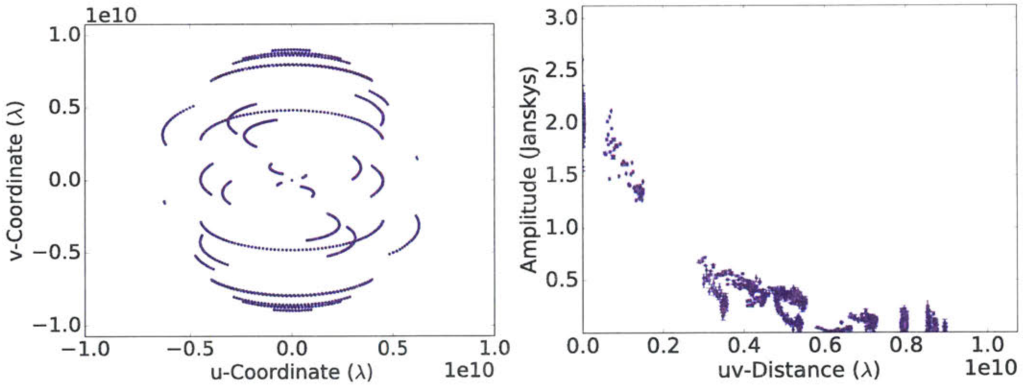
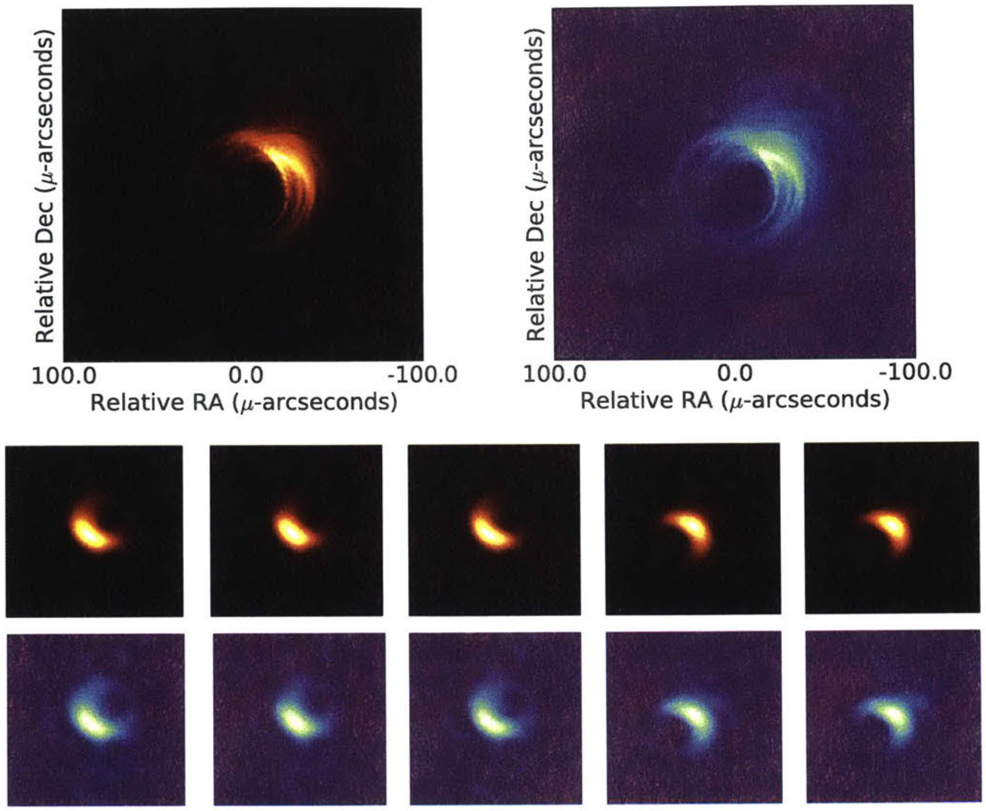


Figure 5.18. Dataset 6: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that the amplitude miscalibration errors causes the amplitude to sometimes exceed the emission's total flux of 2 Jy.

Submissions: Ten submissions were received for this challenge's dataset. The submitted images can be seen in Figure 5.19 for the two colormaps along with the truth image. Interestingly, multiple images were submitted by different participants using the same methods, but with different image structures. The method names and appropriate references for each submission are listed below:



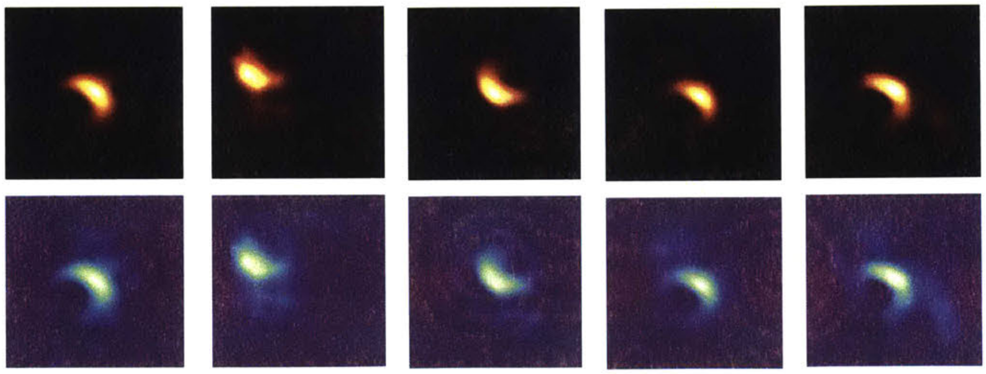
A&T: Amp +Cl-Ph SM1

A&T: Amp-Cal & Amp+Cl-Ph SM 1

A&T: Self-Cal SM 1

A&T: Amp +Cl-Ph SM 2

A&T: Amp-Cal & Amp+Cl-Ph SM 2



A&T: Self-Cal SM 2

C: Amp+Cl-Ph MEMHor

Bo: Amp+ Cl-Ph CHIRP 1

Bo: Amp+ Cl-Ph CHIRP 2

Bo: Bisp CHIRP

Figure 5.19. Dataset 6: Truth (top) and submitted images in two colormaps.

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM 1 & 2) [62]
- Tazaki & Akiyama using Sparse Modeling with Visibility Amplitudes and Closure Phase after Amplitude Calibration (A&T: Amp-Cal& Amp+Cl-Pha SM 1 & 2) [62]
- Tazaki & Akiyama using Sparse Modeling with Self-Calibration (A&T: Self-Cal SM 1 & 2) [62]
- Chael using MEMHorizon with Visibility Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Bouman using CHIRP with Visibility Amplitudes and Closure Phases (Bo: Amp+Cl-Ph CHIRP 1 & 2)
- Bouman using CHIRP with the Bispectrum (Bo: Bisp CHIRP)

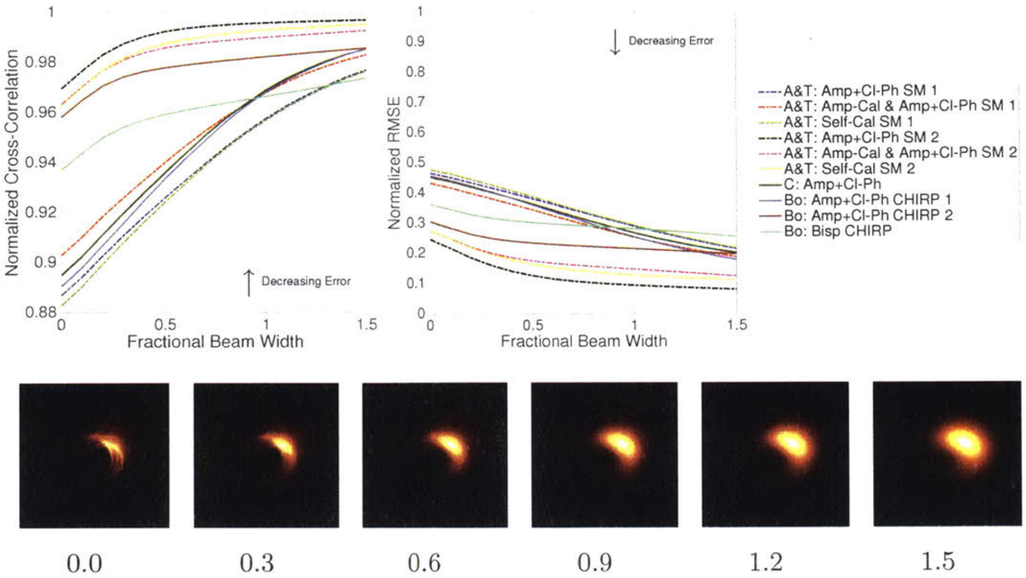


Figure 5.20. Dataset 6: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.20. Multiple participants found different structures that fit the data fairly well. However, only one of these structures was correct, resulting in a bimodal distribution of error. Under these error metrics, A&T: Amp+Cl-Ph SM 2 performed the best.

Qualitative analysis was done by a panel of three judges. The judges were unsure by looking at the data provided which of the two submitted structures was correct, although some guesses were made. We provide selected quotes from each panelist below:

> This appears to be another ring-source with enhanced brightness on the one side (which side looks uncertain). The images with the brightening toward the lower left seem to have a bit simpler overall structure (the others have "extra" emission near the bright edge). So, by Occam's razor one probably should choose the brightening on the lower left. However, if this is a BH accretion disk, many simulations could have "extra" emission near the Doppler boosted edge. So ?? - Mark Reid

> My guess at the underlying true image: This clearly looks like a black hole shadow with familiar decrescient and increscient emissions. -Motoki Kino

> Overall a decent batch of reconstruction, but what is worrying is that the exact intensity distribution over the disk/shadow is very algorithm-dependent. -Fabien Baron

Challenge 3 - Dataset 7

Image: In this dataset a non-astronomical image of a snowman on a black background was used. This image has both large scale (e.g snowballs) and fine scale (e.g. arms) structure. The choice of using a non-astronomical image was made to test the robustness of imaging algorithms to unusual underlying images. The total image field of view is 200 μas and the total flux density is 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA), Arizona (SMT), Mexico (LMT), Chile (ALMA), France (PDB), Spain (PV), and the South Pole (SPT). JCMT and APEX were not included in the array, so there were no short baselines to aid in calibration. Telescopes were assumed to be pointed towards Sgr A*and both thermal noise and atmospheric phase error were included in the visibilities. The uv-coverage for this dataset, along with the visibility amplitude plotted as a function of frequency (uv-distance) is shown in Figure 5.21. In this challenge the total flux of the source was provided to the participants along with the field of view.

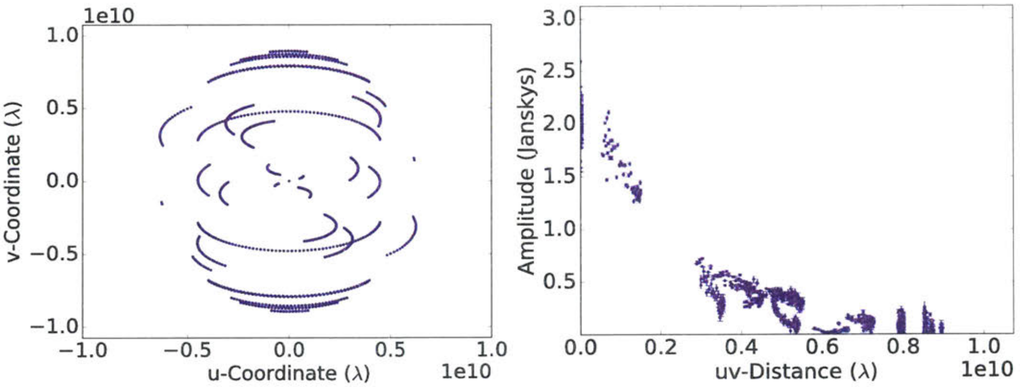
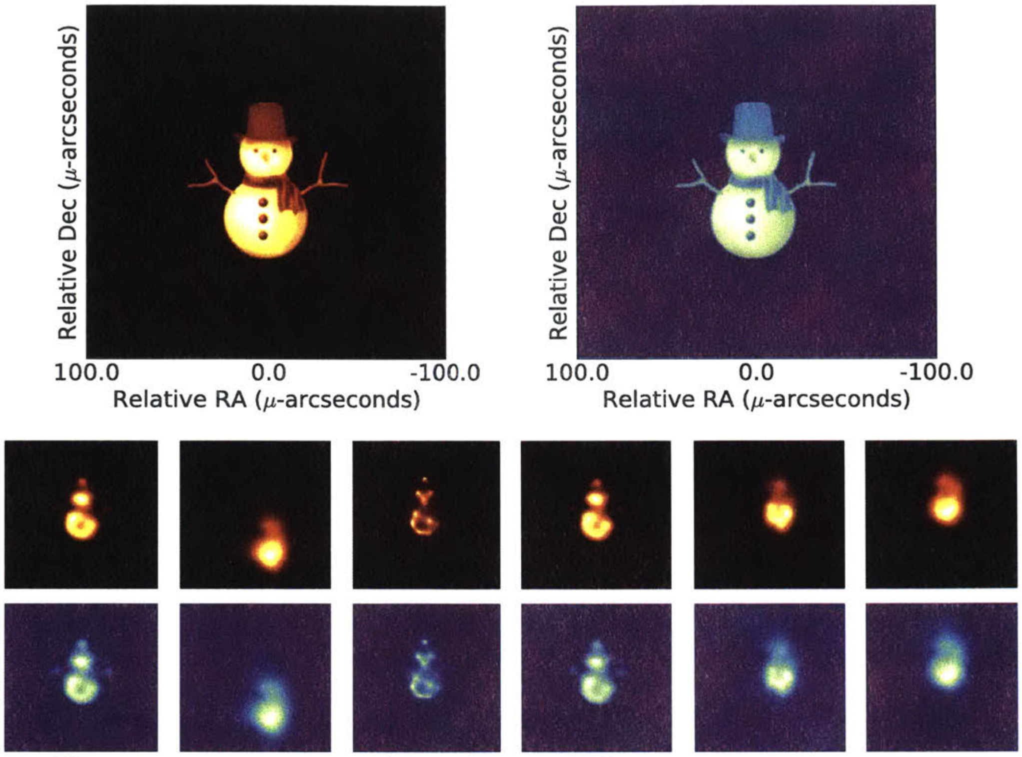


Figure 5.21. Dataset 7: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.

Submissions: Twelve submissions were received for this dataset. The submitted images can be seen in Figure 5.22 for the two colormaps. The method names and appropriate references associated with each of the submissions are listed below:

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: Amp+Cl-Ph SM) [6, 62]
- Akiyama & Tazaki using Sparse Modeling with Self-Calibration (A&T: Self-Cal SM) [6, 62]



A&T: Self-Cal SM

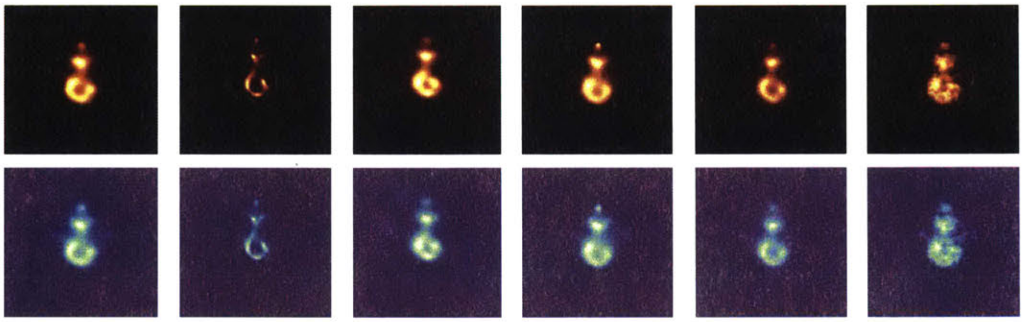
Jor: CLEAN

Joh: Amp+Cl-Ph MEMHor

A&T: Amp +Cl-Ph SM

K&L: CLEAN

M: CLEAN



R: BSMEM

R: MEM

R: MEM Blurred

K&L: BSMEM

Ba: SQUEEZE 1

Ba: SQUEEZE 2

Figure 5.22. Dataset 7: Truth (top) and submitted images in two colormaps.

- Jorstad using CLEAN (Jor: CLEAN) [74]
- Johnson using MEMHor with Visibility Amplitudes and Closure Phases (Joh: Amp+Cl-Ph MEMHor) [26]
- Kim & Lu & Krichbaum using CLEAN (K&L: CLEAN) [115]
- Kim & Lu & Krichbaum using BSMEM (K&L: BSMEM) [24]
- Marscher using CLEAN (M: CLEAN) [74]
- Roelefs using BSMEM (R: BSMEM) [24]
- Roelefs using MEM (R: MEM and R: MEM Blurred) [26]
- Baron using SQUEEZE (Ba: SQUEEZE 1 & 2) [9]

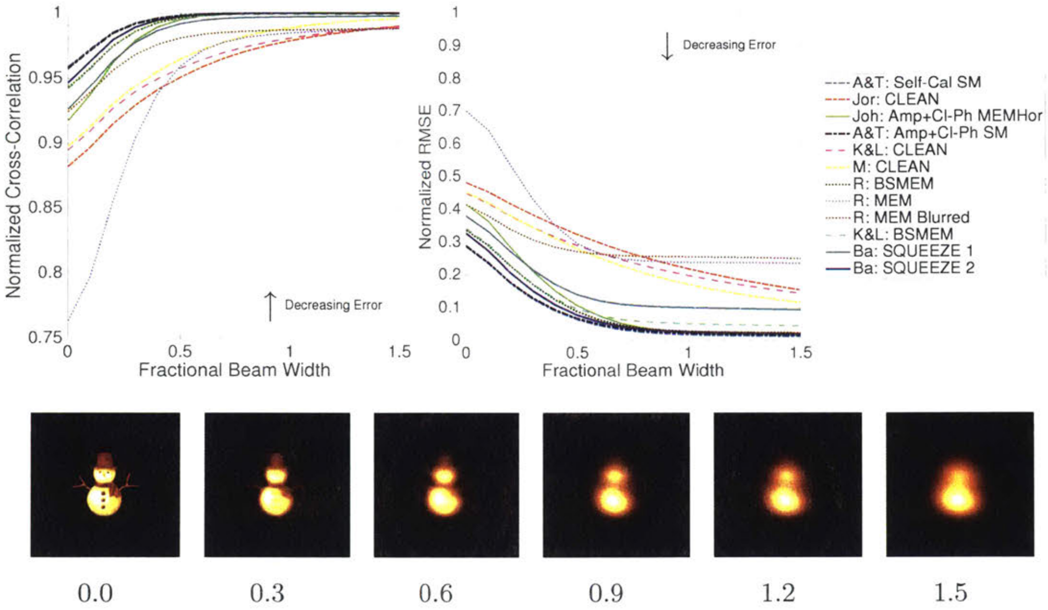


Figure 5.23. Dataset 7: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

Evaluation A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.23. Although most results were very consistent with the true underlying image, A&T: Amp+Cl-Ph SM and A&T: Self-Cal SM performed the best by a slight margin under these error metrics.

Qualitative analysis was done by a panel of two judges. Both judges settled on the two BSMEM (R: BSMEM and K&L: BSMEM) algorithms being among their favorites.

We provide selected quotes from each panelist below:

> These images are consistent with one another, though significant variations exist. Guess at the image: Limb-brightened Frosty the Snowman with his arms open for a hug. - Vincent Fish

> If I were a coauthor on a paper I would feel comfortable endorsing [SparseModeling: Self-Calbrated Visibilities, SparseModeling: Amplutide + Closure Phase, Roelofs: BSMEM], but none of the others (except possibly [Kim: BSMEM]) -Jim Moran

Challenge 4 - Dataset 8

Image: The image used to generate measurements was a black hole image generated by Hung Yi Pu [107] with an additional feature extending towards the top-left added by Mareki Homa. The faint additional feature was added to see how well algorithms could reconstruct a weak feature. The total image field of view was 130 μ-arcseconds and the image's total flux density was rescaled to 2 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA), Arizona (SMT), Mexico (LMT), Chile (ALMA), France (PDB), Spain (PV), and the South Pole (SPT). JCMT and APEX were not included in the array, so there were no short baselines to aid in calibration. The telescopes were assumed to be pointed towards Sgr A*. Thermal noise, atmospheric phase error, and systematic amplitude calibration errors were all included in the visibility measurements. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.24. In this dataset the total flux of the source was provided to the participants, but the field of view was left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size.

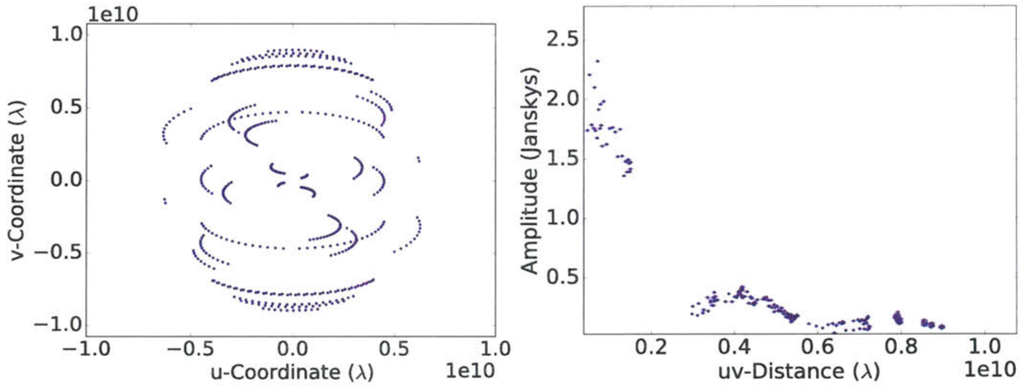
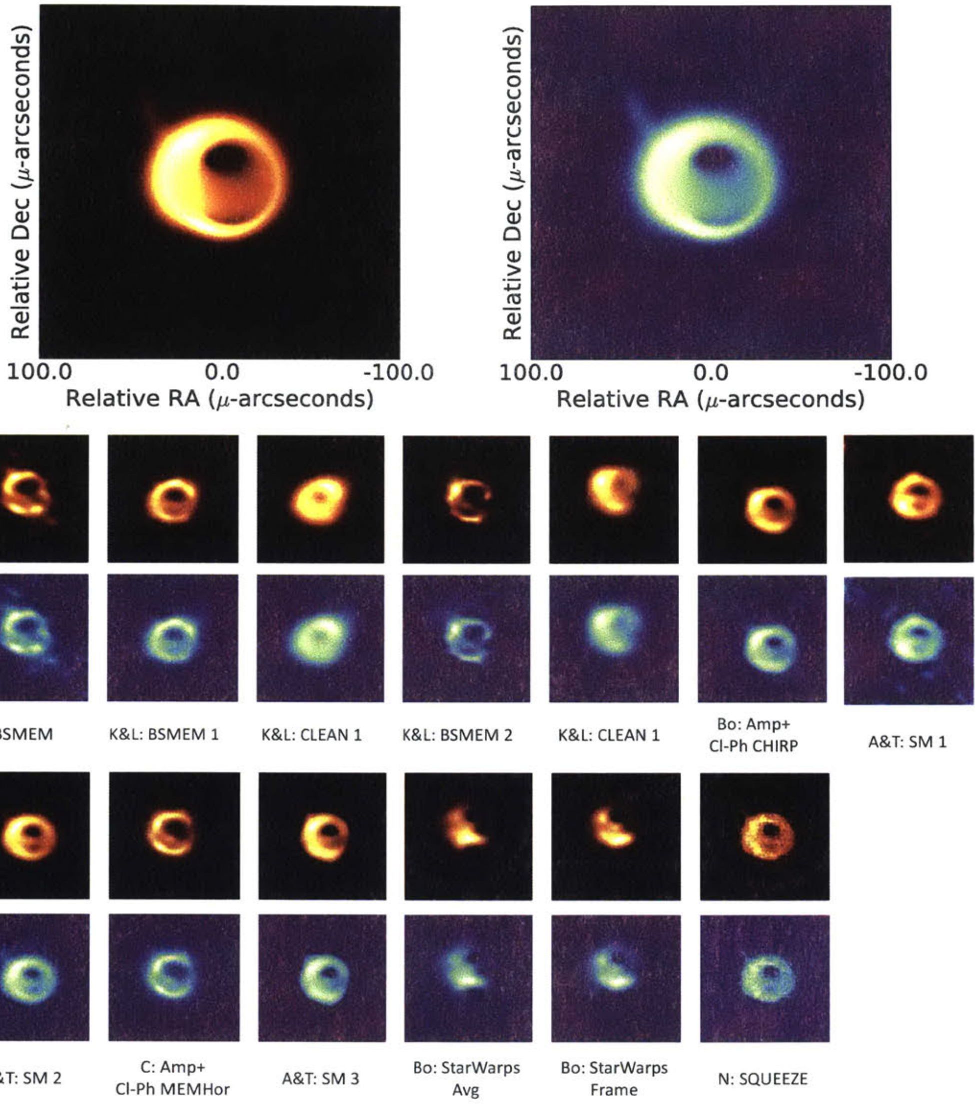


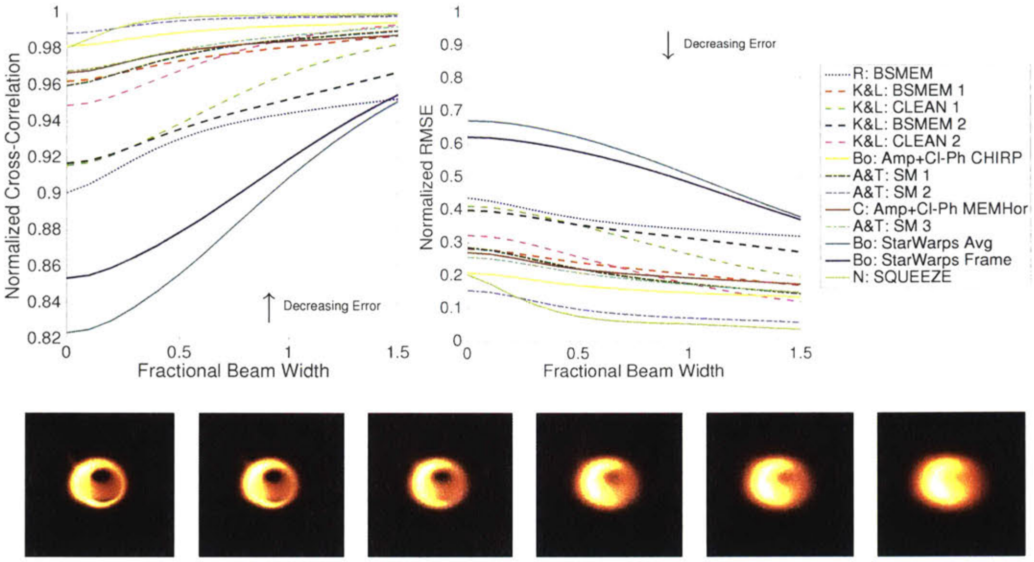
Figure 5.24. Dataset 8: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset. Note that amplitude mis-calibration errors cause the amplitude to sometimes exceed the emission's total flux of 2 Jy.

Submissions: Twelve submissions were received for this dataset. The submitted images can be seen in Figure 5.25 for the two colormaps. The method names and appropriate references associated with each of the submissions are listed below:



R: BSMEM
K&L: BSMEM 1
K&L: CLEAN 1
K&L: BSMEM 2
K&L: CLEAN 1
Bo: Amp+ Cl-Ph CHIRP
A&T: SM 1
A&T: SM 2
C: Amp+ Cl-Ph MEMHor
A&T: SM 3
Bo: StarWarps Avg
Bo: StarWarps Frame
N: SQUEEZE

Figure 5.25. Dataset 8: Truth (top) and submitted images in two colormaps.



0.0 0.3 0.6 0.9 1.2 1.5

Figure 5.26. Dataset 8: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: SM 1 & 2 & 3) [6, 62]
- Chael using MEMHorizon with Visibility Amplitudes and Closure Phases (C: Amp+Cl-Ph MEMHor) [26]
- Kim & Lu & Krichbaum using CLEAN (K&L: CLEAN 1 & 2) [115]
- Kim & Lu & Krichbaum using BSMEM (K&L: BSMEM 1 & 2) [24]
- Roelofs using BSMEM (R: BSMEM) [24]
- Bouman CHIRP using the Visibility Amplitudes and Closure Phases (Bo: Amp+Cl-Ph CHIRP)
- Bouman StarWarps Average using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Avg)
- Bouman StarWarps Frame using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Frame)
- Norris using SQUEEZE (Ba: SQUEEZE 1 & 2) [9]

Evaluation A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.26. Under these metrics, N: SQUEEZE and A&T: SM performed the best, with N:SQUEEZE performing slightly better for matching the structure larger 0.2 of a fractional beam width.

Qualitative analysis was done by a panel of three judges. All judges believed the results were fairly consistent, and that the underlying structure was of a black hole shadow. We provide selected quotes from each panelist below:

> Overall, the structure across most of the images seems to be a black hole silhouette with a possible band running across the middle (disk?). The shadow feature looks to be about 35-40uas in diameter. -Sheperd Doeleman

> ...convey a similar qualitative impression of the source (which does look like a black hole) and (with some yet to be determined method) would likely lead to similar measurements for the size of the shadow of about 40 μ-arcseonds. - Charles Gammie

> From the collection of images it looks pretty convincingly like a black hole, at least to...us who know how a typical simulated GRMHD image looks like. -Heino Falcke

Challenge 4 - Dataset 9

Image: For this dataset a movie, rather than an image, was used to generate visibility measurements. The movie, contributed by Lia Medeiros, was generated by ray-tracing a GRMHD simulation (GRay and HARM) [27, 91]. The black hole model used was SANE (Standard and Normal Evolution) with a spin of a=0.9 and the spin axis oriented horizontally. The image was rescaled such that the size of the black hole shadow is 90% of the expected shadow size for Sgr A*. The movie exhibits considerable variability, including a very large flare that significantly affects the shape of the image. The black hole shadow is fairly visible but the jet is very faint. The emission region of the video is no bigger than 300 μ-arcseconds, and the total flux of the source varies from 1-11 Jy.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), Spain (PV), and the South Pole (SPT). The uv-coverage was generated using a VEX schedule used in the EHT's 2017 observations for Sgr A*. Thermal noise and atmospheric phase error were included in the time-varying measurements. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.27. In this dataset both the varying total flux of the source and field of view were left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size. However, by inspecting the visibility amplitudes on short baselines it could be easily determined that the source was time-varying with a total flux density ranging from 1 to 11 Jy. Sampled video frames from the movie can be seen in Figure 5.30.

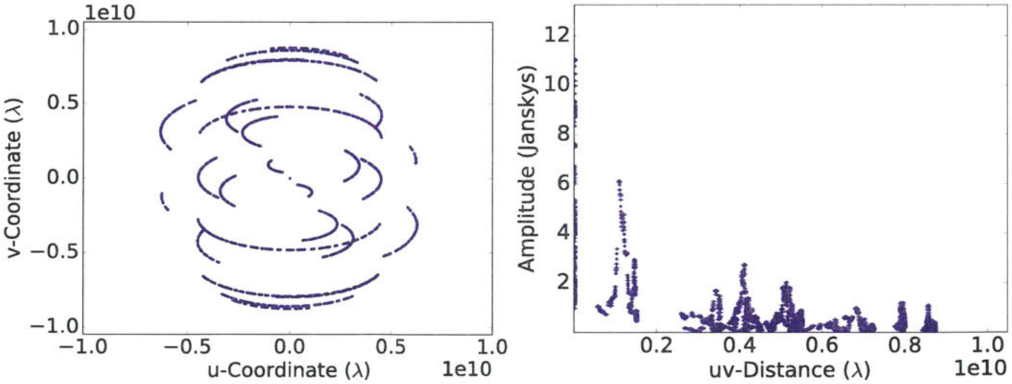
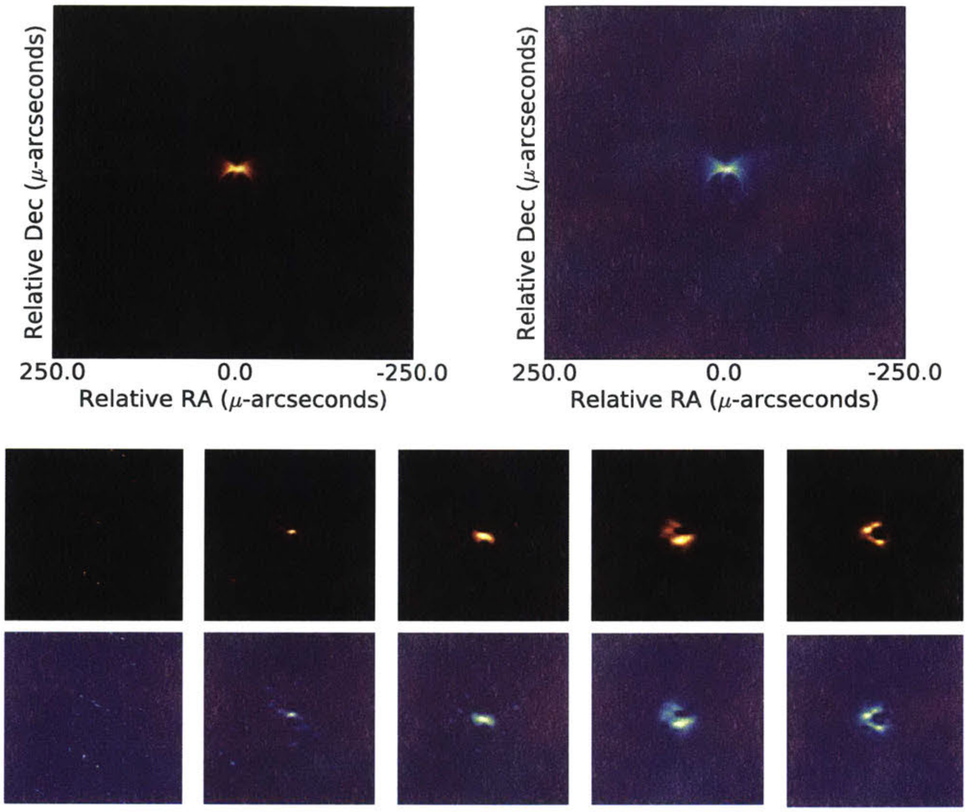
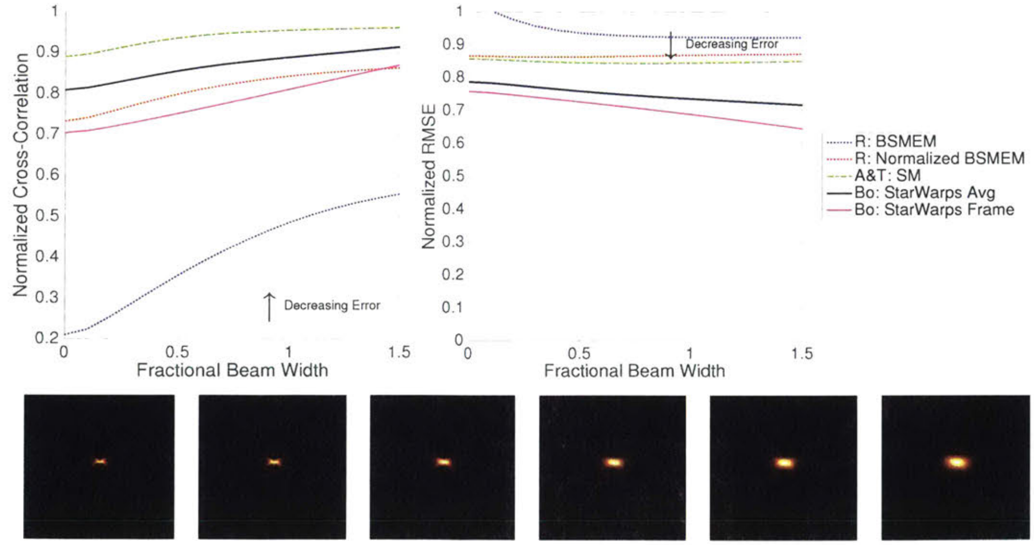


Figure 5.27. Dataset 9: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.



R: BSMEM | R: Normalized BSMEM | A&T: SM | Bo: StarWarps Avg | Bo.: StarWarps Frame

Figure 5.28. Dataset 9: Average truth image (top) and submitted images in two colormaps.



0.0 0.3 0.6 0.9 1.2 1.5

Figure 5.29. Dataset 9: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

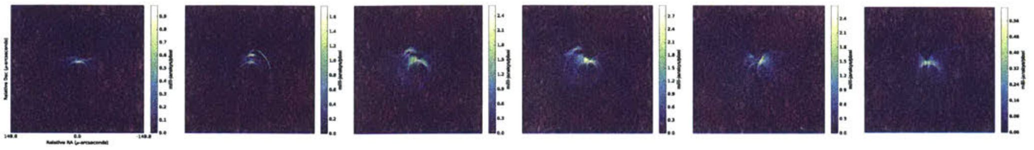


Figure 5.30. Dataset 9: Frames over time from the truth video. Note that there are significant structural changes in the source image over time.

Submissions: Five submissions were received for this dataset. Although the underlying source emission region was time-varying, participants were asked to only submit a single image. The submitted images can be seen in Figure 5.28 for the two colormaps. The method names and appropriate references associated with each of the submissions are listed below:

- Akiyama & Tazaki using Sparse Modeling with Visibility Amplitudes and Closure Phases (A&T: SM) [6, 62]
- Roelofs using BSMEM (R: BSMEM and R: Normalized BSMEM) [24]
- Bouman StarWarps Average using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Avg)
- Bouman StarWarps Frame using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Frame)

Evaluation: A quantitative analysis comparing submissions to the average truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.29. All submitted reconstructions match the true underlying image poorly. A&T: SM matches the best under Normalized Cross Correlation and Bo: StarWarps Frame matches best under Normalized RMSE. This disconnect implies that it is unclear which matches the true image best, and they are both quite different from the true underlying image.

Qualitative analysis was done by a panel of three judges. All judges were unable to make conclusions about the structure of the underlying movie from the submissions. It was also noted that it is unlikely that a single image could represent the underlying structure, as time-variations were especially large in this dataset. We provide selected quotes from each panelist below:

> If these data were collected by the EHT, Id say the following: a) The intrinsic nature of the flux variations could be easily determined by also looking at AGN calibrators. b) Assuming intrinsic structure variations, one would need to attempt dynamical modeling. c) The closure amplitudes should be inspected. -Sheperd Doeleman

> Disappointing set of images, would be difficult to publish any of these unless one were highly favored statistically over the others. -Charles Gammie

> None of the algorithms seem to provide a decent fit anywhere... Id be forced to publish a map by threatening penalty of death -Heino Falcke

Challenge 4 - Dataset 10

Image: For this dataset a movie, rather than an image, was used to generate visibility measurements. The movie, contributed by Lia Medeiros, was generated by ray-tracing a GRMHD simulation (GRay and HARM) [27, 91]. The black hole model used was SANE (Standard and Normal Evolution) with a spin of a=0.7 and the spin axis oriented vertically. This system exhibits considerable variability, but without significant flares. The black hole shadow is fairly prominent but the jet is very faint. The emission region of the video is no bigger than 334 μ-arcseconds, and the total flux of the source varies from 1.5-3.5 Jy. Sampled video frames from the movie can be seen in Figure 5.33.

Data: This dataset was generated using the telescopes in Hawaii (SMA & JCMT), Arizona (SMT), Mexico (LMT), Chile (ALMA & APEX), Spain (PV), and the South Pole (SPT). The uv-coverage was generated using a VEX schedule used in the EHT's 2017 observations for Sgr A*. Thermal noise and atmospheric phase error were included in the time-varying measurements. The uv-coverage and visibility amplitude plotted as a function of frequency (uv-distance) are shown in Figure 5.31. In this dataset both the varying total flux of the source and field of view were left undisclosed. Participants submitted an image of up to 500 μ-arcseconds in size. However, by inspecting the visibility amplitudes on short baselines it could be easily determined that the source was time-varying with a total flux density ranging from 1.5 to 3.5 Jy.

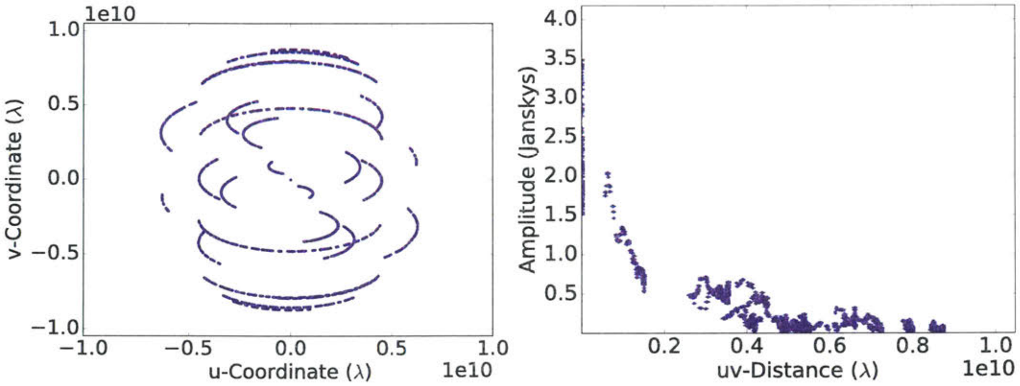
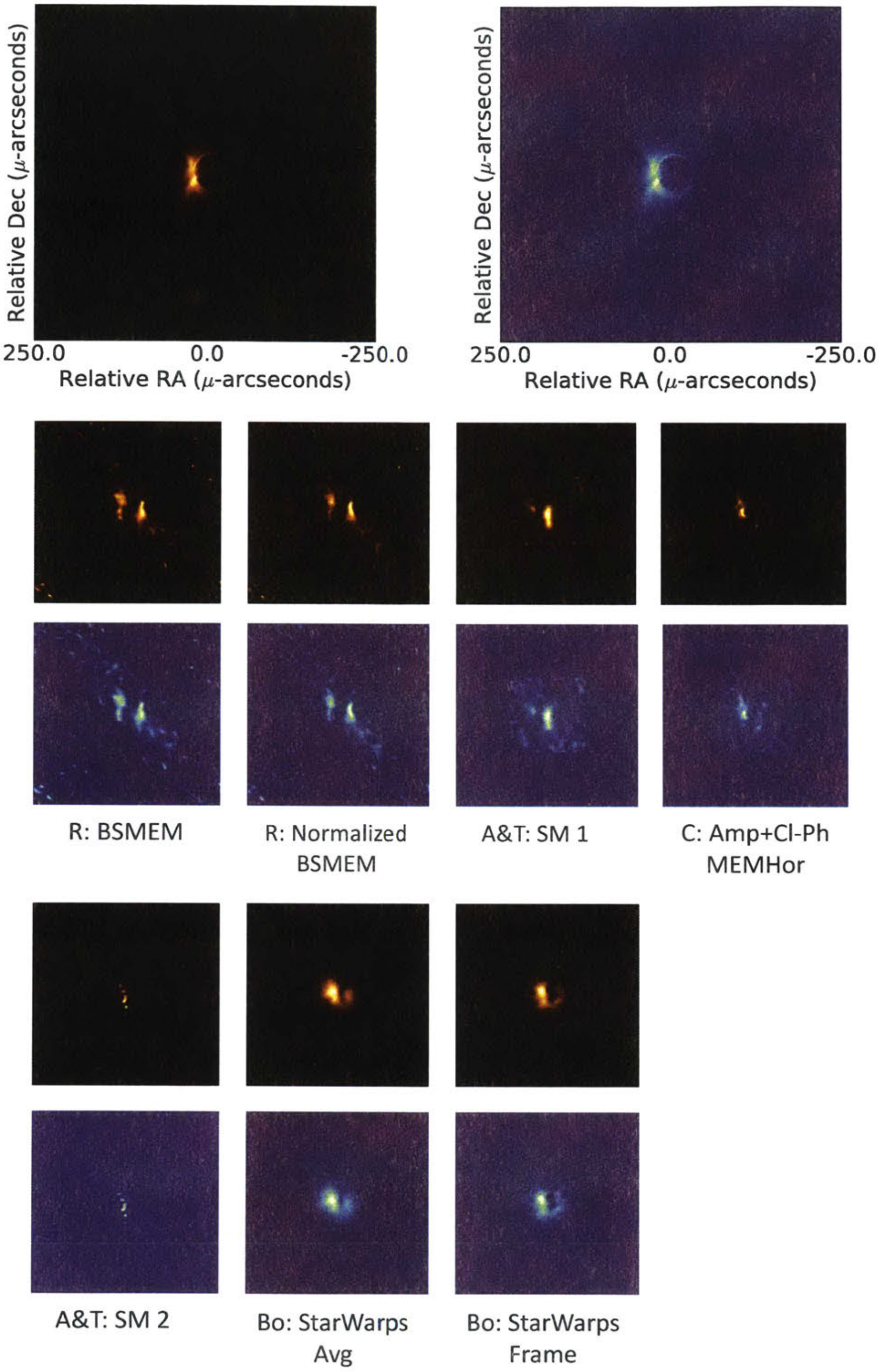


Figure 5.31. Dataset 10: The u-v coverage (left), and amplitude versus u-v distance (right) for the challenge dataset.



R: BSMEM

R: Normalized BSMEM

A&T: SM 1

C: Amp+Cl-Ph MEMHor

A&T: SM 2

Bo: StarWarps Avg

Bo: StarWarps Frame

Figure 5.32. Dataset 10: Average truth (top) and submitted images in two colormaps.

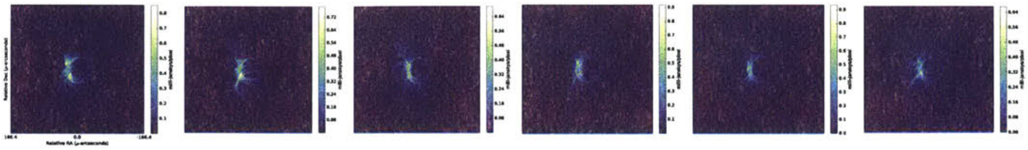
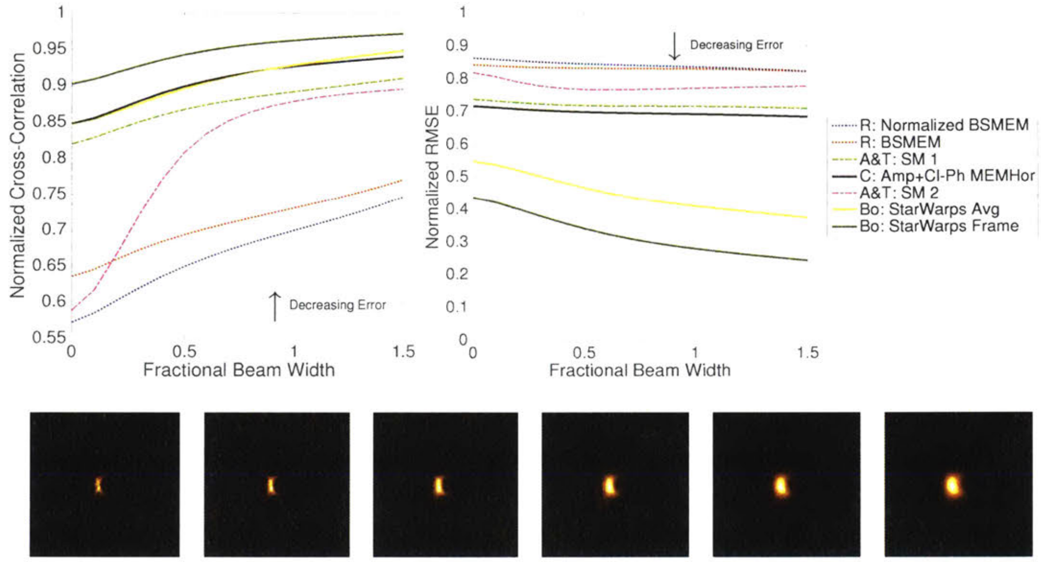


Figure 5.33. Dataset 10: Frames over time from the truth video. Note that there is substantial time-variability, but it the variation is significantly less than in Dataset 9 (see Figure 5.30).



0.0 0.3 0.6 0.9 1.2 1.5

Figure 5.34. Dataset 10: Normalized Cross Correlation and Normalized RMSE for each submission along with a visualization showing the effect of a restoring beam of varying sizes on the truth images.

Submissions: Seven submissions were received for this dataset. Although the underlying source emission region was time-varying, participants were asked to only submit a single image. The submitted images can be seen in Figure 5.32 for the two colormaps. The method names and appropriate references associated with each of the submissions are listed below:

- Tazaki & Akiyama Sparse Modeling using Visibility Amplitudes and Closure Phases (A&T: SM 1 & 2) [62]
- Roelofs using BSMEM (R: BSMEM and R: Normalized BSMEM) [24]

- Chael MEMHorizon using Visibility Amplitudes and Closure Phases (C:Amp+Cl-Ph MEMHor) [26]
- Bouman StarWarps Average using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Avg)
- Bouman StarWarps Frame using the Visibility Amplitudes and Bispectrum (Bo: StarWarps Frame)

Evaluation: A quantitative analysis comparing submissions to the truth image using normalized cross-correlation and normalized RMSE can be seen in Figure 5.34. Under these metrics, Bo: StarWarps Frame performed the best.

Qualitative analysis was done by a panel of three judges. Judges were not confident in their analysis due to the significant variability in results. However, it was noted that a 40 μ-arcsecond structure appears in most reconstructions. We provide selected quotes from each panelist below:

> [R: BSMEM, R: Normalized BSMEM, C: Amp+Cl-Ph MEMHor, Bo: StarWarps Avg, Bo: StarWarps Frame] all appear to have two regions of emission oriented E-W about 40 uas apart (possibly jets?). -Sheperd Doeleman

> there are two conflicting models: .... a black hole with a 40muas shadow [or] an edge-brightened jet. - Charles Gammie

> [Bo: StarWarps] looks a bit suspicious (over-interpreting) but can not be ruled out. -Heino Falcke

5.2.5 Discussion

The first four imaging challenges have been a successful examination of ongoing efforts to develop new imaging techniques that can cope with the unique challenges presented in EHT data. Through these challenges we have not only been able to rigorously evaluate the performance of different methods, but also identify a number of questions that we believe require more attention and discussion before drawing conclusions from images reconstructed with real EHT data.

Encouraging Algorithm Development

The imaging challenges have spurred significant progress in methods being developed for EHT specific data. This progress can be clearly seen by reviewing a single team's submission over time. For instance, Figure 5.35 shows the submissions done using Sparse Modeling by Akiyama & Tazaki over the four challenges. The top row shows the truth image, and the bottom row a result submitted by the team. In the first challenge, although results were somewhat consistent with the truth image, they contained a lot of spurious artifacts. However, after learning about their method's weaknesses through the challenges, the team was able to refine their method and produce some of the best submissions by the third and fourth challenges.

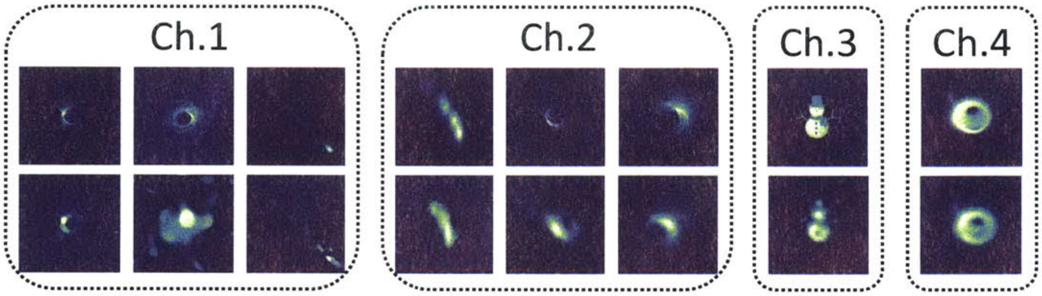


Figure 5.35. Images submitted to the first four imaging challenges by Akiyama & Tazaki using Sparse Modeling. The truth image is shown on the top and the submitted image right below. Note the substantial improvement of their method in recovering the correct structure of a static source.

Breaking Algorithms: By generating challenge datasets that are sparser, noisier, or just outside of the algorithm's comfort zone, participants are often introduced to flaws in their method. Exposing weaknesses of imaging methods in this way has lead to the most substantial algorithm development.

One example of this can be seen with the large field of view required for imaging in dataset 3 of the first challenge. Previously, imaging algorithms for the EHT were solely tested on examples with small fields of view. Thus, most algorithms did not perform well on data with extended structure (see Figure 5.10). However, in the time between the first and second challenges, participants improved their methods based on these findings, resulting in much better performance on the large field of view example in dataset 4 (see Figure 5.13).

To encourage progress, we continue to introduce increasingly realistic datasets. In the fourth, and most recent, challenge we introduced time-variability into datasets 9 and 10 - generating visibilities from a movie rather than a still image. This change violated most imaging method's fundamental assumptions, and resulted in poor results (see Figures 5.28 and 5.32). Even methods designed to work with time-varying data (e.g. StarWarps) had trouble, and did not recover an accurate image for the 9th dataset (this dataset contained much more variability than is normally tested for StarWaprs). We hope that the introduction of increasingly realistic datasets, such as these, will continue to progress development of imaging methods.

Improving Data Constraints: The progress of some methods (e.g. Bo: CHIRP and C: MEMHor) can also be attributed to comparisons done through the imaging challenge. For instance, results of the first challenge highlighted the importance of a method's choice of constraining data products. An example of this is in Chael's three submissions (C: Closure Only MEMHor, C: Amp+Cl-Ph MEMHor, and C: Bisp MEMHor). These reconstructions were obtained using the same optimization strategy and image priors, but the data products that were constrained in imaging were changed, leading to drastically different results (see Figures 5.7 and 5.10). In particular, Chael's submissions for the first challenge that constrained both the visibility amplitudes and closure phases consistently performed well across all datasets.

Figure 5.36 shows results of dataset 3 obtained by constraining the visibility amplitude and closure phases versus just the bispectrum. The clear success of the amplitude and closure phase imaging suggested that additional study into algorithms that utilize the visibility amplitudes (instead of the bispectrum) may be fruitful. In subsequent challenges, Bo: CHIRP and C: MEMHor improved their submissions by submitting results that were obtained by constraining visibility amplitudes. The clear advantage of this choice can be seen by comparing Bo: Amp+Cl-Ph CHIRP and Bo: Bisp CHIRP in challenge 2's datasets (see Figures 5.13 and 5.16). Adding these additional visibility constraints has proved to be especially helpful in large field of view imaging.

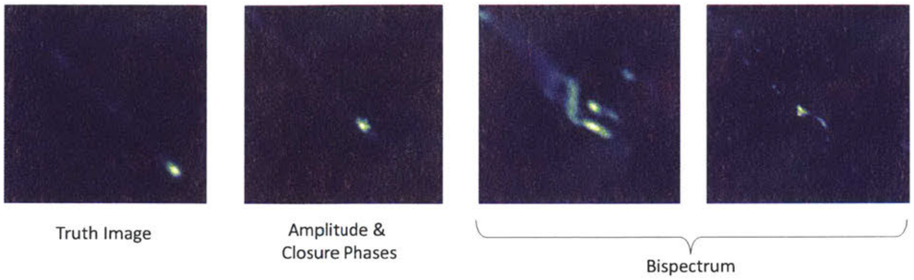


Figure 5.36. For extended emissions that require a large field of view during imaging, the imaging challenges helped to identify that it was very important to constrain the visibility amplitudes. As the bispectrum is the product of three terms, there can be a lot of uncertainty in reconstructing each visibility's amplitude when imaging purely with the bispectrum. This can result in poor reconstructions, especially when the field of view is not well constrained.

Identifying Biases

One of the initial motivations for setting up the imaging challenges was to help in understanding the biases of different algorithms. After analyzing the results of the first four challenges some conclusions can be made.

As expected, Bayesian-style methods outperform traditional CLEAN approaches on EHT data (see Figures 5.22 and 5.25). These Bayesian-style methods also generally achieve some level of super-resolution, although at different degrees. For instance, MEMHorizon results in the most super-resolved recovered images. MEMHorizon's recovered images can be helpful in identifying structures smaller than the nominal beam width (see Figure 5.7), and thus are often favored by judges. However the resulting images by MEMHorizon are generally sparser than the true underlying image. CHIRP on the other hand usually produces blurrier results than MEMHorizon. Empirically, CHIRP results generally become consistent with the truth image at around 1/2 the nominal beam width.

Dataset 7 of challenge 3 used a picture of a snowman as the true underlying image. This dataset was designed to test methods' ability to recover the correct structure in the case of an unusual or surprising underlying truth image. For the most part the imaging algorithms produced an image that accurately reflects the basic structure of the snowman (see Figure 5.23). Results like these give us confidence that we are not overly biasing our methods to produce images of a black hole shadow.

Identifying Degeneracies

As image reconstruction from VLBI data is an ill-posed problem, there are an infinite number of possible solutions that perfectly explain the data. Nevertheless, it is generally the case that most imaging methods converge to a very similar looking image. However, the level of degeneracy in EHT data is still an open question: are there multiple reasonable image structures that fit the data equally well, and if so, when does this happen?

Dataset 6 is a surprising example of when a serious degeneracy exists in EHT quality data. In this dataset, two sets of structures were identified as being consistent with the data provided. Figure 5.37 shows these two structures along with plots comparing the true and measured amplitude and closure phases. These resulting images appeared to be flips of one another, but surprisingly, both fit the measured data equally well. From the provided data, both the imaging participants and expert judges were unable to decide which structure was most likely to be correct.

Upon inspection, it was concluded that one possible reason for this degeneracy is due to the presence of noisy closure phases close to zero. The visibilities of a flipped image will simply be conjugated with respect to the original image. Thus, if an image has small closure phases originally, the flipped image will also have a small closure phase. Consequently, a large amount of error on a small closure phase could result in uncertainty in the direction of the image's structure.

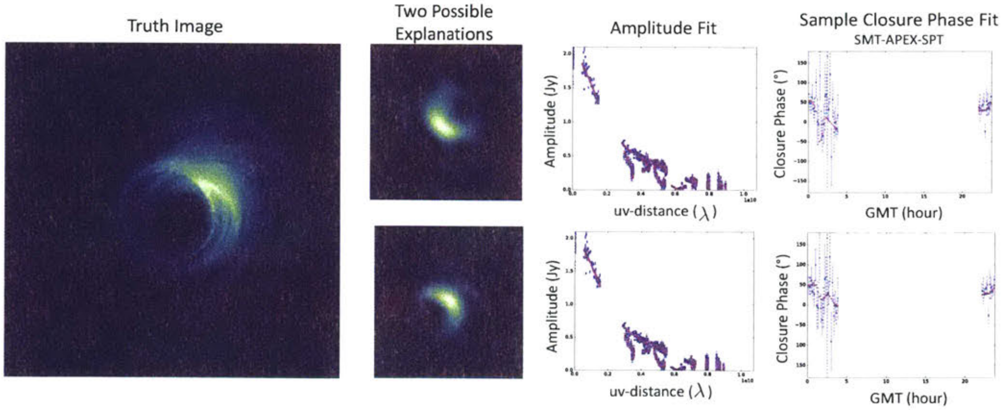


Figure 5.37. A case of degeneracy in EHT data. Two possible explanations fit the simulated EHT data equally well. In most cases the closure phases appeared nearly identical, and only small differences existed (that fit well within the error expected on measurements). We believe this degeneracy exists due to small closure phases with large error.

This example brings to our attention a flaw with current imaging algorithms: most algorithms just produce a single final image. However, the collected EHT data may have serious degeneracies, and it is important that our algorithms have a way of finding them. Currently, finding different solutions requires the user to adjust parameters and manually search the space of possible images. Instead, we should be developing imaging methods that alert us to the level of degeneracy in the data by finding all reasonable images that fit the data well.

Identifying Masquerading Structures

For the second dataset, most algorithms reproduced the primary image features. However, the prominent ring feature in this image was not caused by a black hole shadow, but instead a jet funnel wall emission. See Figure 5.38. This caused the expert judges to incorrectly believe there existed a black hole shadow with a smaller diameter of 30 μ-arcseconds.

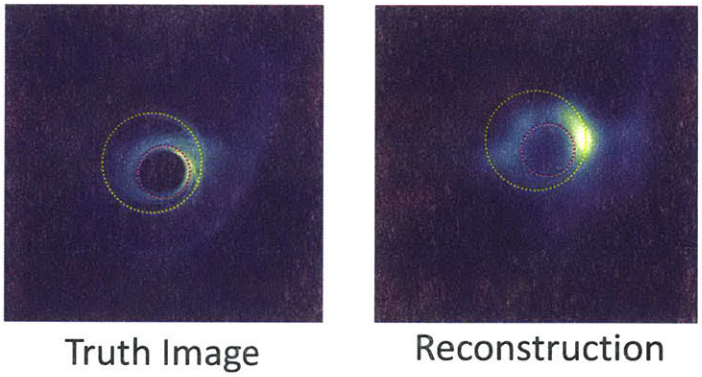


Figure 5.38. The black hole's lensed photon ring is highlighted by the yellow ring. However, a funnel wall causes a more prominent ring feature that shows up in the reconstruction, highlighted in red. This example emphasizes the importance of further study into features that may masquerade as a black hole photon ring.

This result is especially interesting, considering the results presented in [36]. In [36] a smaller primary emission region of 37 μ-arcseconds lead the authors to believe that it is most likely that we are observing Sgr A* edge-on, with a Doppler boosted edge contributing to a smaller bright region than would be produced by a face-on black hole. However, perhaps the primary emission region is not due to the black hole's photon ring, but instead should be attributed to another source of emission.

Dataset 2 emphasizes the importance of further study into features that may masquerade as a black hole shadow. It is important for us to understand what situations could cause similar features to appear, and in those cases, determine if there ways to distinguish them from a black hole's photon ring.

Beginning a Conversation About Confidence

All algorithms make different imaging assumptions, and thus produce different images, even in the best cases. However, despite these differences, there is often some consensus as to the primary structure in the image. In fact, comparing independently reconstructed images resulting from different methods can often help in identifying which structures are most likely to be real and which are spurious artifacts.

Figure 5.39. Five submissions reconstructed using the same data but using different imaging algorithms.

Are the images shown in Figure 5.39, originating from dataset 5, consistent with one another? Which structures are real and which are probably just artifacts? One observation made through the judging panels is that different people have different standards on how similar images must look to believe an interpretation of them. Thus, it is imperative that a conversation is started now on what constitutes consistency between images. Perhaps, quantitative measures should also be developed to help determine the consistency between a set of images. Addressing these questions early will help avoid discord when deciding if a reconstructed image from EHT data should be released.

6

Turning Corners into Cameras

THE ability to see around obstructions would prove valuable in a wide range of applications. As just two examples, remotely sensing occupants in a room would be valuable in search and rescue operations, and the ability to detect hidden, oncoming vehicles and/or pedestrians would be valuable in collision avoidance systems [16]. Although often not visible to the naked eye, in many environments, light from obscured portions of a scene is scattered over many of the observable surfaces. This reflected light can be used to recover information about the hidden scene (see Fig. 6.1). In this chapter, we exploit the vertical edge at the corner of a wall to construct a "camera" that sees beyond the wall. Since vertical wall edges are ubiquitous, such cameras can be found in many environments.

The radiance emanating from the ground in front of a corner, e.g., at the base of a building, is influenced by many factors: the albedo, shape, and BRDF of its surface, as well as the light coming from the full hemisphere above it. Assuming the ground has a significant diffuse component, a majority of the reflected light comes from the surroundings that are easily seen from the observer's position next to the occluding wall (the visible region is shaded in yellow in Fig. 6.1a). However, emitted and reflected light from behind the corner, hidden from the observer, also has a small effect on the ground's radiance in the form of a subtle gradient of light encircling the corner; this is not a shadow, but is instead what is referred to as a *penumbra.*

The faint penumbra on the ground is caused by the reflection of an increasing amount of light from the hidden scene. To illustrate this, imagine standing with your shoulder up against the building's wall (refer to the leftmost picture of Fig. 6.1b). At this position you are unable to see any of the scene behind the corner. However, as you slowly move away from the wall, walking along the magenta circle shown in Fig. 6.1a, you see an increasing amount of the scene. Eventually, the hidden scene comes fully into view. Similarly, different points on the ground reflect light integrated from differently-sized

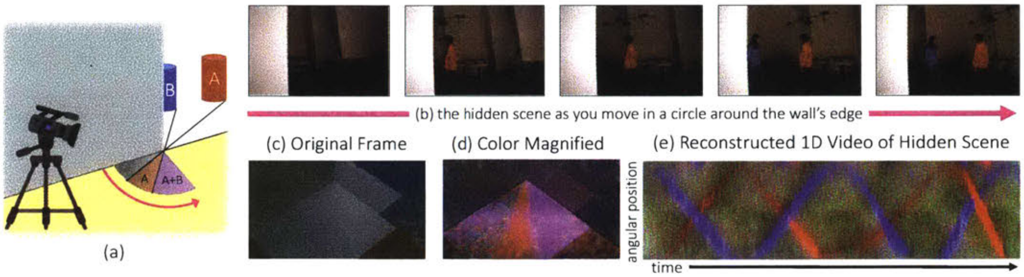


Figure 6.1. We construct a 1-D video of an obscured scene using RGB video taken with a consumer camera. The stylized diagram in (a) shows a typical scenario: two people—one wearing red and the other blue—are hidden from the camera's view by a wall. Only the region shaded in yellow is visible to the camera. To an observer walking around the occluding edge (along the magenta arrow), light from different parts of the hidden scene becomes visible at different angles (see sequence (b)). Ultimately, this scene information is captured in the intensity and color of light reflected from the corresponding patch of ground near the corner. Although these subtle irradiance variations are invisible to the naked eye (c), they can be extracted and interpreted from a camera position from which the entire obscured scene is hidden from view. Image (d) visualizes these subtle variations in the highlighted corner region. We use temporal frames of these radiance variations on the ground to construct a 1-D video of motion evolution in the hidden scene. Specifically, (e) shows the trajectories over time that specify the angular position of hidden red and blue subjects illuminated by a diffuse light.

fractions of the hidden scene.

Now imagine someone has entered the hidden portion of the scene. This person would introduce a small change to the light coming from an angular slice of the room. From behind the corner this change would often not be perceptible to the naked eye. However, it would result in a subtle change to the penumbra; see Fig. 6.1c and (d). We use these subtle changes, recorded from standard video cameras, to construct a 1-D version of how the hidden scene beyond the corner evolves with time; see Fig. 6.1e.

Section 6.1 summarizes related work that puts the present contribution in context. Section 6.2 shows how, using our proposed methods, it is possible to identify the number and location of people in a hidden scene. Section 6.3 develops how parallax created by a pair of adjacent edges, such as in a doorway, can be used to triangulate the 2D position of moving people over time. Experimental results are shown for a number of indoor and outdoor environments with varied flooring, including carpet, tile, hardwood, concrete, and brick.

■ 6.1 Related Work

In this section we describe previous non-line-of-sight (NLoS) methods. Previous methods used to see past or through occluders have ranged from using WiFi signals [4] to exploiting random specular surfaces [41, 137]. In this summary, we emphasize a few active and passive approaches that have previously been used to see past occluders and image hidden scenes.

Recovery under Active Illumination: Past approaches to see around corners have largely involved using time-of-flight (ToF) cameras [56, 82, 117, 135]. These methods involve using a laser to illuminate a point that is visible to both the observable and hidden scene, and measuring how long it takes for the light to return [119, 135]. By measuring the light's time of flight, one can infer the distance to objects in the hidden scene, and by measuring the light's intensity, one can often learn about the reflectance and curvature of the objects [116]. Past work has used ToF methods to infer the location [75], size and motion [46, 101], and shape [129] of objects in the hidden scene. These methods have also been used to count hidden people [134].

ToF cameras work well in estimating the depths of hidden objects, however, they have some limitations. First, they require specialized and comparatively expensive detectors with fine temporal resolution. Second, they are limited in how much light they can introduce in the scene to support imaging. Third, they are vulnerable to interference from ambient outdoor illumination. By contrast, our proposed real-time passive technique operates in unpredictable indoor and outdoor environments with inexpensive consumer cameras, without additional illumination.

In [79] a laser is used to indirectly illuminate an object behind an occluder. Using a standard camera the authors are then able to identify the position of the hidden object. Similar to our proposed work, [79] uses a standard camera; however, their proposed system has a number of limitations. Namely, they require controlled conditions where the geometry of the unknown moving object is rigid, and its shape and material are either known or can be closely modeled by a single oriented surface element. In contrast, our method requires minimal prior information, is completely passive, and has been shown to work in many natural settings.

Passive Recovery: Other work has previously considered the possibility of using structures naturally present in the real world as cameras. Naturally occurring pinholes (such as windows) or pinspecks have been previously used for non-line-of-sight imaging [28, 125]. In addition, specular reflections off of human eyes have been used to image hidden scenes [99]. Although these accidental cameras can be used to reconstruct 2-D images, they require a more specialized accidental camera scenario than the simple edges we propose to use in this work.

The technique presented in [133] also detects and visualizes small, often imperceptible, color changes in video. However, in this work, rather than just visualize these tiny color changes, we interpret them in order to reconstruct a video of a hidden scene.

■ 6.2 Edge Cameras

An edge camera system consists of four components: the visible and hidden scenes, the occluding edge, and the ground, which reflects light from both scenes. We refer to the (ground) plane perpendicular to the occluding edge as the *observation plane.* By analyzing subtle variations in the penumbra at the base of an edge, we are able to deduce a hidden subject's pattern of motion.

The reflected light from a surface at point p, with normal $\hat{n}$, is a function of the incoming light L_i' as well as the surface's albedo a and BRDF β. Specifically,

$$L_o'(p, \hat{v}_o) = a(p) \int L_i'(p, \hat{v}_i)\, \beta(\hat{v}_i, \hat{v}_o, \hat{n})\, \gamma(\hat{v}_i, \hat{n})\, \mathrm{d}\hat{v}_i, \tag{6.1}$$

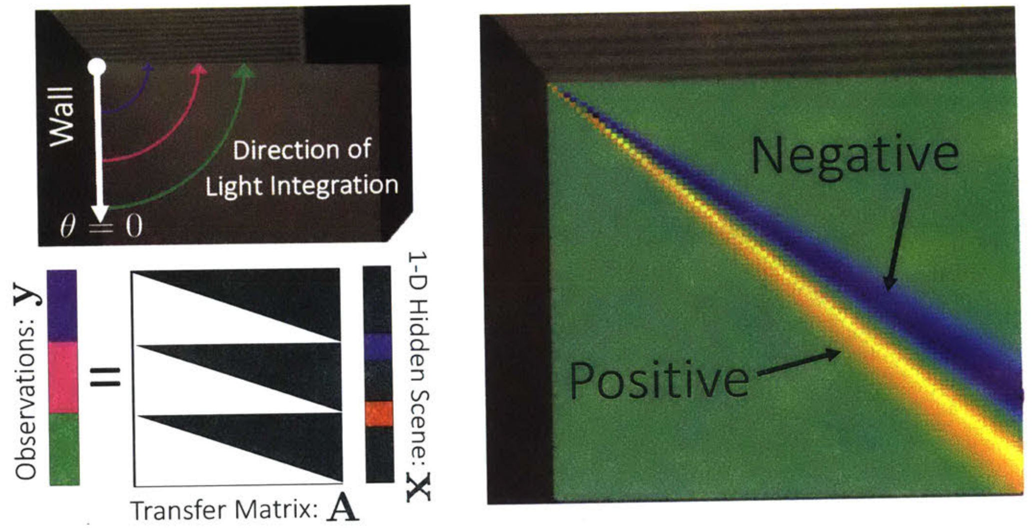
where $\hat{v}_i$ and $\hat{v}_o$ denote the incoming and outgoing unit vectors of light at position $p = (r, \theta)$, respectively, and $\gamma(\hat{v}_i, \hat{n}) = \hat{v}_i \cdot \hat{n}$. We parameterize p in polar coordinates, with the origin centered at the occluding edge and $\theta = 0$ corresponding to the angle parallel to the wall coming from the corner (refer to Fig. 6.2). For simplicity, we assume the observation plane is Lambertian, and that the visible and hidden scene are modeled as light emitted from a large celestial sphere, parameterized by right ascension α and declination δ. Under these assumptions, we simplify (6.1):

$$L_o'(r, \theta) = a(r, \theta) \int_{\alpha=0}^{2\pi} \int_{\delta=0}^{\pi/2} L_i(\alpha, \delta)\, \mathrm{d}\alpha\, \mathrm{d}\delta, \tag{6.2}$$

where $L_i = L_i'\gamma$. Furthermore, since the occluding edge blocks light from $[\pi + \theta, 2\pi]$ at the radial line θ,

$$L_o'(r, \theta) = a(r, \theta) \left[L_v + \int_{\phi=0}^{\theta} L_h(\phi)\, \mathrm{d}\phi \right], \tag{6.3}$$

for $L_v = \int_{\alpha=0}^{\pi} \int_{\delta=0}^{\pi/2} L_i(\alpha, \delta)\, \mathrm{d}\alpha\, \mathrm{d}\delta$ and $L_h(\phi) = \int_{\delta=0}^{\pi/2} L_i(\pi + \phi, \delta)\, \mathrm{d}\delta$. By inspecting (6.3) we



(a) Constructing Transfer Matrix (b) Sample Estimation Gain Image

Figure 6.2. In (a), the transfer matrix, **A**, is shown for a toy situation in which observations lie along circles around the edge. In this case, **A** would simply be a repeated lower triangular matrix. (b) contains an example estimation gain image, which describes the matrix operation performed on observations $\mathbf{y}^{(t)}$ to estimate $\mathbf{x}^{(t)}$. As predicted, the image indicates that we are essentially performing an angular derivative in recovering a frame of the 1-D video.

can see that the intensity of light on the penumbra is explained by a constant term, L_v, which is the contribution due to light visible to the observer (shaded in yellow in Fig. 6.1a), and a varying angle dependent term which integrates the light in the hidden scene, L_h. For instance, a radial line at $\theta = 0$ only integrates the light from the scene visible to the observer, while the radial line $\theta = \pi/2$ reflects the integral of light over the entire visible and hidden scenes.

Then, if we assume that $\frac{d}{d\theta} a(r, \theta) \approx 0$[1], the derivative of the observed penumbra recovers the 1-D angular projection of the hidden scene:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} L_o'(r, \theta) \approx a(r, \theta)\, L_h(\theta). \tag{6.4}$$

But what happens if someone walks into the hidden scene at time t, changing $L_h^0(\theta)$

---

[1] In practice, we subtract a background frame to substantially remove per-pixel albedo variations. Refer to Section 6.2.1

to $L_h^t(\theta)$? In this case, the spatial derivative of the temporal difference encodes the angular change in lighting:

$$\frac{d}{d\theta}\left[L_o'^t(r,\theta) - L_o'^0(r,\theta)\right] \approx a(r,\theta)\left[L_h^t(\theta) - L_h^0(\theta)\right] \tag{6.5}$$

In other words, the angular derivative of the penumbra's difference from the reference frame is a signal that indicates the angular change in the hidden scene over time. In practice, we obtain good results assuming $a(r,\theta) = 1$ and using the cameras' native encoded intensity values while subtracting the temporal mean as a background frame (see Section 6.2.1).

6.2.1 Method

Using a video recording of the observation plane, we generate a 1-D video indicating the changes in a hidden scene over time. These 1-D angular projections of the hidden scene, viewed over many time-steps, reveal the trajectory of a moving object behind the occluding edge.

Likelihood: At each time t, we relate the observed M-pixels on the projection plane, $\mathbf{y}^{(t)}$, to the 1-D angular projection of the hidden scene, $L_h^{(t)}(\theta)$. We formulate a discrete approximation to our edge camera system by describing the continuous image $L_h^{(t)}(\theta)$ using N terms, $\mathbf{x}^{(t)}$. The observations $\mathbf{y}^{(t)}$ then relate to the unknown parameters $\mathbf{x}^{(t)}$ and $L_v^{(t)}$ by a linear matrix operation:

$$\mathbf{y}^{(t)} = L_v^{(t)} + \mathbf{A}\mathbf{x}^{(t)} + \mathbf{w}^{(t)}, \qquad \mathbf{w}^{(t)} \sim \mathcal{N}(0, \lambda^2 \mathbb{1}),$$

where the $M \times N$ matrix $\mathbf{A}$ is defined by the geometry of the system. More explicitly, each row m of $\mathbf{A}$ integrates the portion of the hidden scene visible from observation m, $\mathbf{y}_m^{(t)}$. In the simplified case of observations that lie on a circle around the occluding edge, $\mathbf{A}$ would simply be a constant lower-triangular matrix; see Fig. 6.2a.

Let $\widetilde{\mathbf{A}}$ be the column augmented matrix $[\mathbf{1}\ \mathbf{A}]$. We can then express the likelihood of an observation given $\mathbf{x}^{(t)}$ and $L_v^{(t)}$ as:

$$p(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, L_v^{(t)}) = \mathcal{N}\left(\widetilde{\mathbf{A}}\left[L_v^{(t)}\ \mathbf{x}^{(t)T}\right]^T, \lambda^2 \mathbb{1}\right). \tag{6.6}$$

Prior: The signal we are trying to extract is very small relative to the total light intensity on the observation plane. Therefore, to improve the quality of results, we enforce spatial smoothness of $\mathbf{x}^{(t)}$. We use a simple L2 smoothness regularization over adjacent parameters in $\mathbf{x}^{(t)}$. This corresponds, for a gradient matrix $\mathbf{G}$, to using the prior

$$\begin{aligned} p(\mathbf{x}^{(t)}) \propto \prod_{n=1}^{N-1} \exp\left[-\frac{1}{2\sigma_1^2}\|\mathbf{x}^{(t)}[n] - \mathbf{x}^{(t)}[n-1]\|_2^2\right] & \\ \prod_{n=1}^{N} \exp\left[-\frac{1}{2\sigma_2^2}\|\mathbf{x}^{(t)}[n]\|_2^2\right] & \qquad (6.7) \\ = \mathcal{N}(\mathbf{0}, \sigma_1^2(\mathbf{G}^T\mathbf{G})^{-1} + \sigma_2^2\mathbb{1}). & \qquad (6.8) \end{aligned}$$

Inference: We seek a maximum a posteriori (MAP) estimate of the hidden image coefficients, $\mathbf{x}^{(t)}$, given M observations, $\mathbf{y}^{(t)}$, measured by the camera. By combining the defined Gaussian likelihood and prior distributions, we obtain a Gaussian posterior distribution of $\mathbf{x}^{(t)}$ and $L_v^{(t)}$,

$$\begin{aligned} p(\mathbf{x}^{(t)}, L_v^{(t)}|\mathbf{y}^{(t)}) &= \mathcal{N}\left(\left[\hat{L}_v^{(t)} \ \hat{\mathbf{x}}^{(t)T}\right]^T, \Sigma^{(t)}\right) \\ \Sigma^{(t)} &= \left[\lambda^{-2}\widetilde{\mathbf{A}}^T\widetilde{\mathbf{A}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{G}^T\mathbf{G}}{\sigma_1^2} + \frac{\mathbb{1}}{\sigma_2^2} \end{bmatrix}\right]^{-1} \\ \left[\hat{L}_v^{(t)} \ \hat{\mathbf{x}}^{(t)T}\right]^T &= \Sigma^{(t)}\lambda^{-2}\widetilde{\mathbf{A}}^T\mathbf{y}^{(t)}, \qquad (6.9) \end{aligned}$$

where the maximum a posteriori estimate is given by $\hat{\mathbf{x}}^{(t)}$.

To better understand the operation that is being performed to obtain the 1-D reconstruction, we visualize each row of the matrix $\Sigma^{(t)}\lambda^{-2}\widetilde{\mathbf{A}}^T$. We refer to each reshaped row of this matrix as the *estimation gain image*. An example estimation gain image is shown in Fig. 6.2b. As expected, the matrix operation is computing an angular derivative over the observation plane. Note that although earlier we assumed $\frac{d}{d\theta}a(r,\theta) \approx 0$, in reality the albedo simply needs to be orthogonal to the zero-mean pie-wedges in each estimation gain image. We expect violations from this assumption to be small.

Implementation Details

Rectification: All of our analysis thus far has assumed we are observing the floor parallel to the occluding edge. However, in most situations, the camera will be observing the

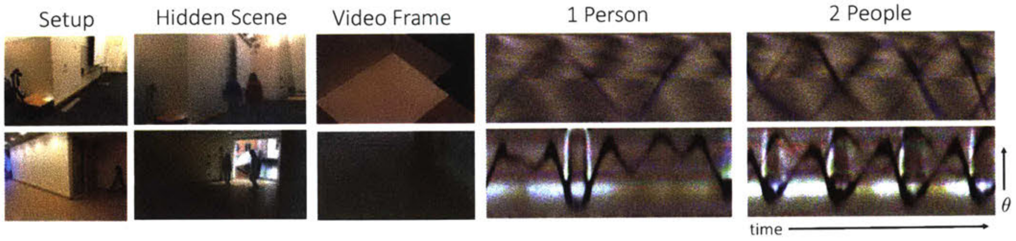


Figure 6.3. One-dimensional reconstructed videos of indoor, hidden scenes. Results are shown as space-time images for sequences where one or two people were walking behind the corner. In these reconstructions, the angular position of a person, as well as the number of people, can be clearly identified. Bright vertical line artifacts are caused by additional shadows appearing on the penumbra. We believe horizontal line artifacts result from sampling on a square grid.

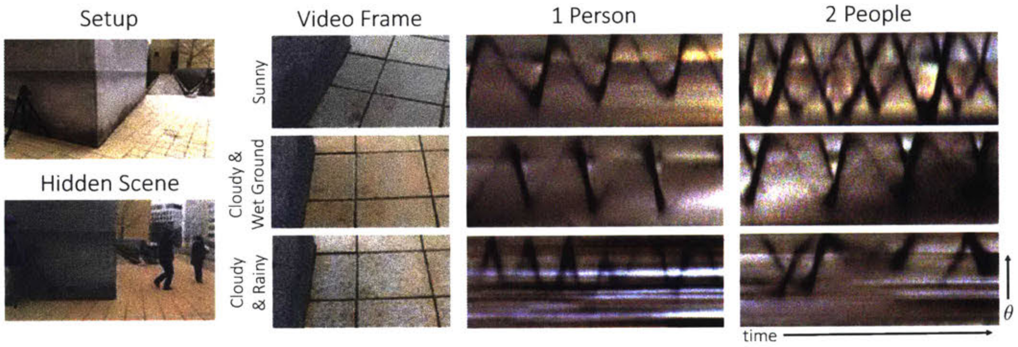
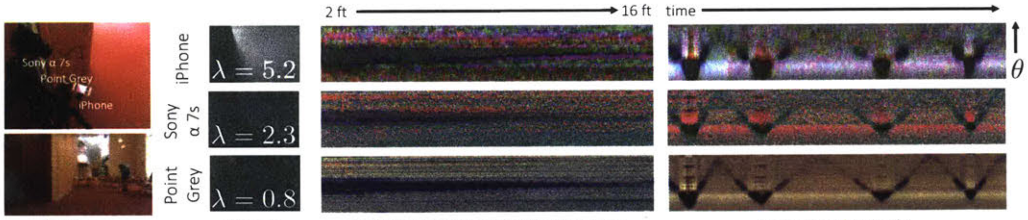


Figure 6.4. 1-D reconstructed videos of a common outdoor, hidden scene under various weather conditions. Results are shown as space-time images. The last row shows results from sequences taken while it was beginning to rain. Although artifacts appear due to the appearing raindrops, motion trajectories can be identified in all reconstructions.

projection plane at an angle. In order to make the construction of the matrix $\mathbf{A}$ easier, we begin by rectifying our images using a homography. In these results, we assume the ground is perpendicular to the occluding edge, and estimate the homography using either a calibration grid or regular patterns, such as tiles, that naturally appear on the ground. Alternatively, a known camera calibration could be used.

Background Subtraction: Since we are interested in identifying temporal differences in a hidden scene due to a moving subject, we must remove the effect of the scene's background illumination. Although this could be accomplished by first subtracting a background frame, L_o^0, taken without the subject, we avoid requiring the availability of such a frame. Instead, we assume the subject's motion is roughly uniform over the video, and use the video's mean image in lieu of a true background frame. We found



(a) Setup (b) Noise (c) Walking from 2 to 16 feet at a 45° Angle (d) Walking Randomly

Figure 6.5. The result of using different cameras on the reconstruction of the same sequence in an indoor setting. Three different 8-bit cameras (an iPhone 5s, a Sony Alpha 7s, and an uncompressed RGB Point Grey) simultaneously recorded the carpeted floor. Each camera introduced a different level of sensor noise. The estimated standard deviation of per-pixel sensor noise, λ, is shown in (b). We compare the quality of two sequences in (c) and (d). In (c), we have reconstructed a video from a sequence of a single person walking directly away from the corner from 2 to 16 feet at a 45 degree angle from the occluded wall. This experiment helps to illustrate how signal strength varies with distance from the corner. In (d), we have done a reconstruction of a single person walking in a random pattern. In (c) the hidden person does not change in angular position. Thus, for these results, we subtract an average background frame computed from a different portion of the video sequence.

that in sequences containing people moving naturally, background subtraction using the average video frame worked well.

Temporal Smoothness: In addition to spatial smoothness we could also impose temporal smoothness on our MAP estimate. $\hat{\mathbf{x}}^{(t)}$. This helps to further regularize our result, at the cost of some temporal blurring. However, to emphasize the coherence among results, we do not impose this additional constraint. Each 1-D image, $\mathbf{x}^{(t)}$, that we show is independently computed. Results obtained with temporal smoothness constraints are shown in Appendix C.

Parameter Selection: The noise parameter λ^2 is set for each video as the median variance of estimated sensor noise. The regularization parameters σ_1 and σ_2 are empirically set to 0.1 for all results.

■ 6.2.2 Experiments and Results

Our algorithm reconstructs a 1-D video of a hidden scene from behind an occluding edge, allowing users to track the motions of obscured, moving objects. In all results shown, the subject was not visible to an observer at the camera.

We present results as space-time images. These images contain curves that indicate the angular trajectories of moving people. All results, unless specified otherwise, were generated from standard, compressed video taken with a SLR camera.

Environments

We show several applications of our algorithm in various indoor and outdoor environments. For each environment, we show the reconstructions obtained when one or two people were moving in the hidden scene.

Indoor: In Fig. 6.1e we show a result obtained from a video recorded in a mostly dark room. A large diffuse light illuminated two hidden subjects wearing red and blue clothing. As the subjects walked around the room, their clothing reflected light, allowing us to reconstruct a 1-D video of colored trajectories. As correctly reflected in our reconstructed video, the subject in blue occludes the subject in red three times before the subject in red becomes the occluder.

Fig. 6.3 shows additional examples of 1-D videos recovered from indoor edge cameras. In these sequences, the environment was well-lit. The subjects occluded the bright ambient light, resulting in the reconstruction's dark trajectory. Note that in all the reconstructions, it is possible to count the number of people in the hidden scene, and to recover important information such as their angular position and speed, and the characteristics of their motion.

Outdoor: In Fig. 6.4 we show the results of a number of videos taken at a common outdoor location, but in different weather conditions. The top sequences were recorded during a sunny day, while the bottom two sequences were recorded while it was cloudy. Additionally, in the bottom sequence, raindrops appeared on the ground *during* recording, while in the middle sequence the ground was fully saturated with water. Although the raindrops cause artifacts in the reconstructed space-time images, you can still discern the trajectory of people hidden behind the wall.

Video Quality:

In all experiments shown thus far we have used standard, compressed video captured using a consumer camera. However, video compression can create large, correlated noise that may affect our signal. We have explored the effect video quality has on results. To do this, we filmed a common scene using 3 different cameras: an iPhone 5s, a Sony Alpha 7s, and a uncompressed RGB Point Grey. Fig. 6.5 shows the results of this experiment assuming different levels of i.i.d. noise. Each resulting 1-D image was reconstructed from a single frame. The cell phone camera's compressed videos resulted in the noisiest reconstructions, but even those results still capture key features of the subject's path.

Velocity Estimation

The derivative of a person's trajectory over time, $\theta^{(t)}$, indicates their angular velocity. Fig. 6.6 shows an example of the estimated angular velocity obtained from a single edge camera when the hidden subject was walking roughly in a circle. Note that the person's angular size and speed are both larger when the person is closer to the corner. Such cues can help approximate the subject's 2-D position over time.

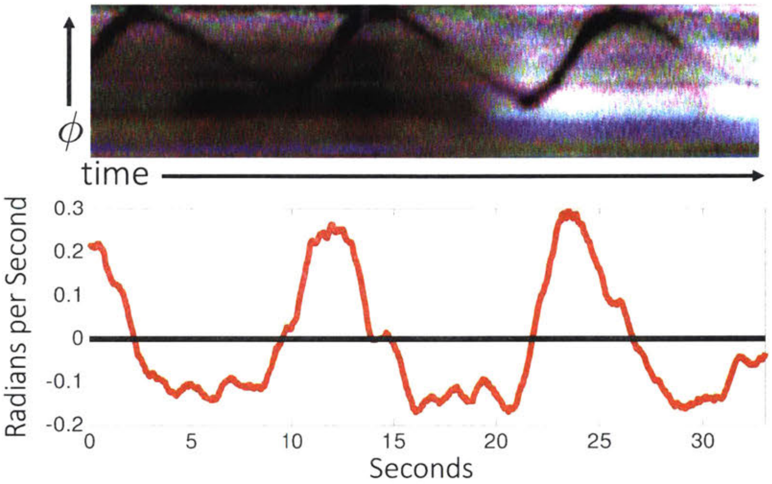


Figure 6.6. A subject's reconstructed angular velocity relative to the corner as a function of time. In this sequence, a person was walking in circles far from the corner.

6.3 Stereo Edge Cameras

Although the width of a track recovered in the method of the previous section can give some indication of a hidden person's relative range, more accurate methods are possible by exploiting adjacent walls. For example, when a hidden scene is behind a doorway, the pair of vertical doorway wall edges yield a pair of corner cameras. By treating the observation plane at the base of each edge as a camera, we can obtain stereo 1-D images that we can then use to triangulate the absolute position of a subject over time.

6.3.1 Method

A single edge camera allows us to reconstruct a 90° angular image of an occluded scene. We now consider a system composed of four edge cameras, such as an open doorway, as

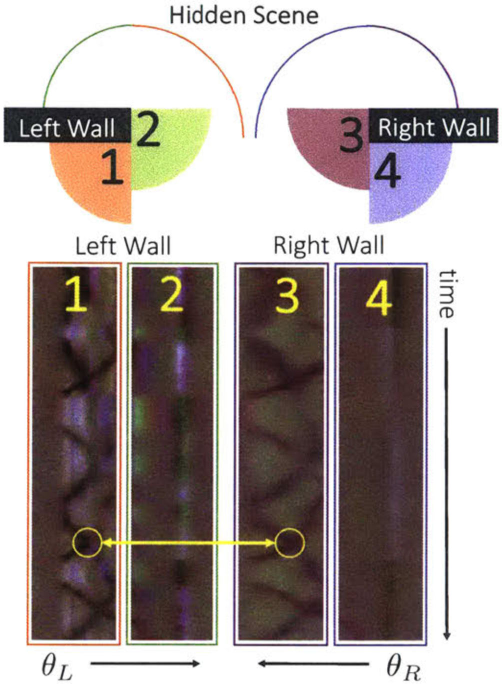


Figure 6.7. The four edges of a doorway contain penumbras that can be used to reconstruct a 180° view of a hidden scene. The top diagram indicates the penumbras and the corresponding region they describe. Parallax occurs in the reconstructions from the left and right wall. This can be seen in the bottom reconstruction of two people hidden behind a doorway. Numbers/colors indicate the penumbras used for each 90° space-time image.

illustrated in Fig. 6.7. Each side of the doorway contains two adjacent edge cameras, whose reconstructions together create a 180° view of the hidden scene.

The two sides of the doorway provide two views of the same hidden scene, but from different positions. This causes an offset in the projected angular position of the same person (see Fig. 6.8). Our aim is to use this angular parallax to triangulate the location of a hidden person over time. Assume we are observing the base of a doorway, with walls of width w separated by a distance B. A hidden person will introduce an intensity change on the left and right wall penumbras at angles of $\theta_L^{(t)}$ and $\theta_R^{(t)}$, respectively. From this correspondence, we can triangulate their 2-D location.

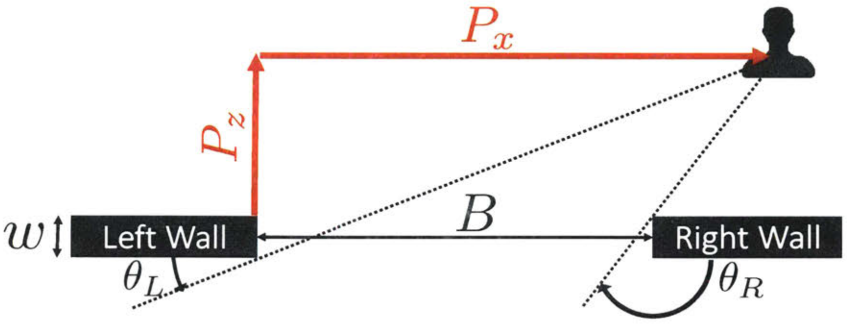


Figure 6.8. A hidden person will introduce an intensity change on the left and right wall penumbras at angles of $\theta_L^{(t)}$ and $\theta_R^{(t)}$, respectively. Once these angles have been identified, we can recover the hidden person's two-dimensional location using Eq. 6.11.

$$P_z^{(t)} = \frac{B - \eta^{(t)}}{\cot\theta_L^{(t)} + \cot\theta_R^{(t)}} \tag{6.10}$$

$$P_x^{(t)} = P_z^{(t)} \cot\theta_L^{(t)} \tag{6.11}$$

$$\eta^{(t)} = \begin{cases} w\cot(\theta_R) & P_x \le 0 \\ 0 & 0 \le P_x \le B \\ w\cot(\theta_L) & P_x \ge B \end{cases} \tag{6.12}$$

where (P_x, P_z) are the x- and z-coordinate of the person. We define the top corner of the left doorway, corner 1 in Fig. 6.7, as $(P_x, P_z) = (0, 0)$.

Assuming the wall is sufficiently thin compared to the depth of moving objects in the hidden scene, the $\eta^{(t)}$ term can be ignored. In this case, the relative position of the person can be reconstructed without any knowledge of the absolute geometry of the doorway (e.g. B or w). In all results shown in this paper, we have made this assumption.

Identifying Trajectories: While automatic contour tracing methods exist [77], for simplicity, in our stereo results, we identify the trajectories of objects in the hidden scene manually by tracing a path on the reconstructed space-time images.

■ 6.3.2 Experiments and Results

We demonstrate the ability of our method to localize the two-dimensional position of a hidden object using four edge cameras, such as in a doorway. We present a series of experiments in both controlled and uncontrolled settings. Full sequences, indicating the ground truth motions, and additional results can be found in Appendix C.

Controlled Environment: To demonstrate the ability to infer depth from stereo edge cameras we constructed a controlled experiment. A monitor displaying a slowly moving green line was placed behind two walls, separated by a baseline of 20 cm, at a distance of roughly 23, 40, 60, and 84 cm. Fig. 6.10b shows sample space-time reconstructions of each 180° edge camera. The depth of the green line was then estimated from manually identified trajectories obtained from these space-time images. Empirically estimated error ellipses are shown in red for a subset of the depth estimates.

Natural Environment: Fig. 6.9 shows the results of estimating 2-D positions from doorways in natural environments. The hidden scene consists of a single person walking in a circular pattern behind the doorway. Although our reconstructions capture the cyclic nature of the subject's movements, they are sensitive to error in the estimated trajectories. Refer to Section 6.3.3. Ellipses indicating empirically estimated error have been drawn around a subset of the points.

■ 6.3.3 Error Analysis

There are multiple sources of error that can introduce biases into location estimates. Namely, inaccuracy in localizing the projected trajectories, and mis-calibration of the scene cause error in the estimates. We discuss the effects of some of these errors below. Further derivations and analysis can be seen in Appendix C.

Trajectory Localization: Because P_z scales inversely with $cot(\theta_L)+cot(\theta_R)$, small errors in the estimated projected angles of the person in the left and right may cause large errors in the estimated position of the hidden person, particularly at larger depths. Assuming Gaussian uncertainty in the left and right angular trajectories, σ_{θ_L} and σ_{θ_R}, the uncertainty in the estimated position of the hidden person will not be Gaussian. However, the standard deviation of empirical distributions through sampling, as seen in Figs. 6.9 and 6.10, can be informative. Additionally, by using standard error propagation of independent variables, we can compute a first order approximation of the uncertainty. For instance, the uncertainty in the z position, σ_{P_z}, is

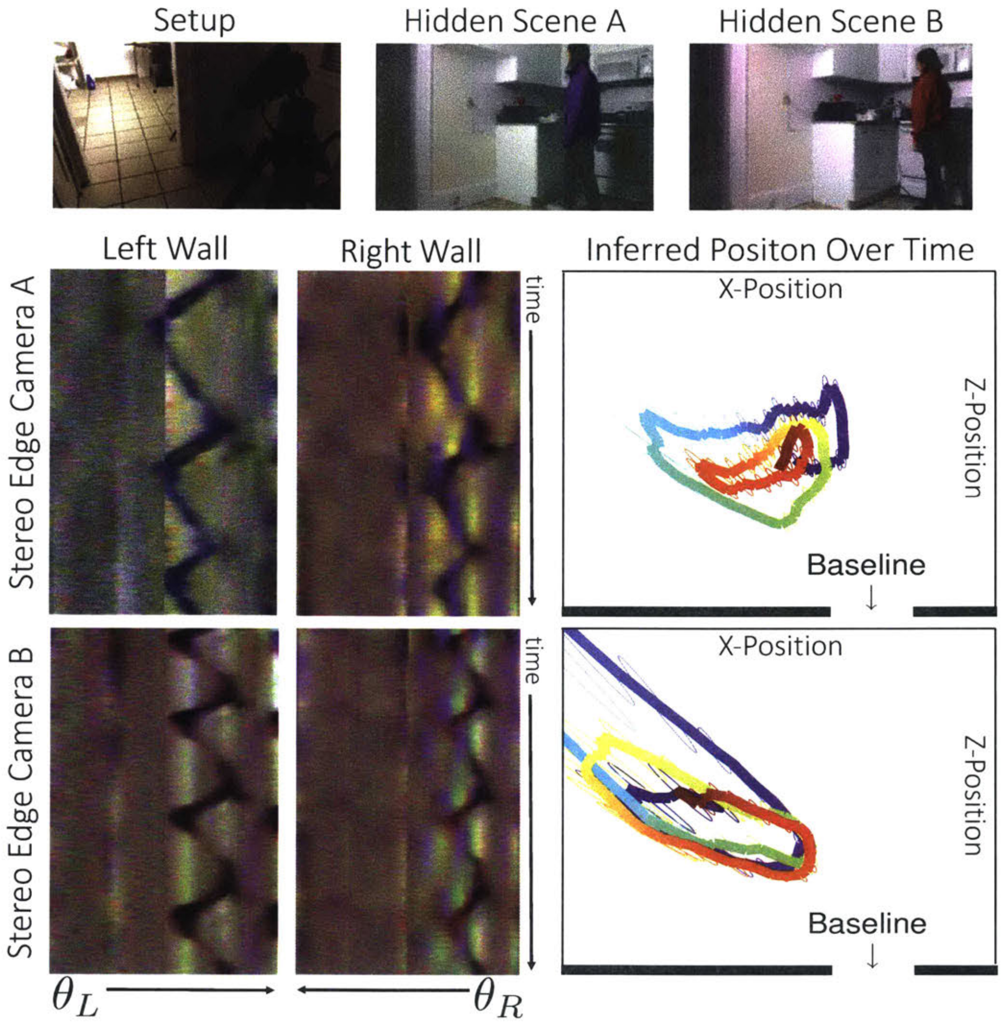
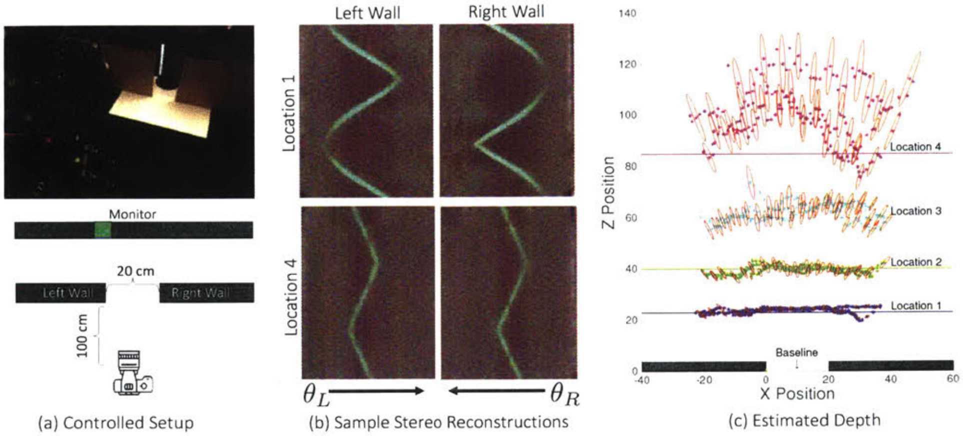


Figure 6.9. The results of our stereo experiments in a natural setting. Each sequence consists of a single person walking in a roughly circular pattern behind a doorway. The 2-D inferred locations over time are shown as a line from blue to red. Error bars indicating one standard deviation of error have been drawn around a subset of the points. Our inferred depths capture the hidden subject's cyclic motion, but are currently subject to large error. A subset of B's inferred 2-D locations have been cut out of this figure.

$$\sigma_{P_z} = B\sqrt{\frac{\sigma_{\theta_L}^2 \csc^4\theta_L + \sigma_{\theta_R}^2 \csc^4\theta_R}{(\cot\theta_L + \cot\theta_R)^4}} \tag{6.13}$$



(a) Controlled Setup (b) Sample Stereo Reconstructions (c) Estimated Depth

Figure 6.10. Controlled experiments were performed to demonstrate the ability to infer depth from stereo edge cameras. A monitor displaying a moving green line was placed behind an artificial doorway (a) at four locations corresponding to 23, 40, 60, and 84 cm, respectively. (b) shows sample reconstructions done of the edge cameras for the left and right wall when the monitor was placed at 23 and 84 cm. Using tracks obtained from these reconstructions, the 2-D position of the green line in each sequence was estimated over time (c). The inferred position is plotted with empirically computed error ellipses (indicating one standard deviation of noise).

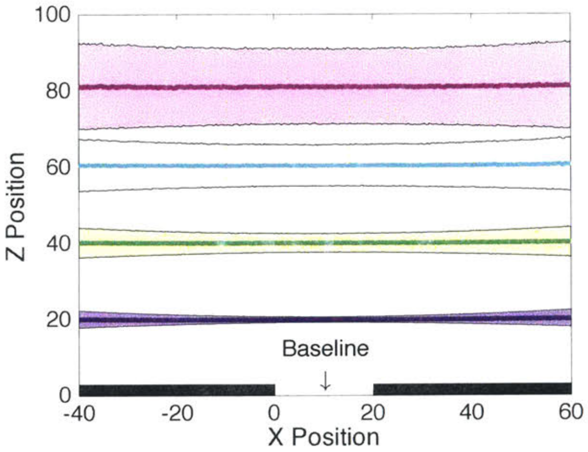


Figure 6.11. The empirical means plus or minus one standard deviation of the estimated P_z as a function of its x-coordinate, assuming true P_z of 20, 40, 60, and 80. Here, the two corner location errors at each of the boundaries of the doorway are independent and subject to $\sigma^2_{\Delta x} = \sigma^2_{\Delta z} = 0.04$.

Corner Identification: Misidentifying the corner of each occluding edge will cause systematic error to the estimated 2-D position. To determine how erroneously identifying a corner affects our results, we consider the following situation: a doorway of baseline $B = 20$ obscuring a bright object at angular position θ in an otherwise dark scene.

Assuming the offset from the true corner location is drawn from an independent Gaussian distribution, we can calculate the error between the estimated and true angular position, and then subsequently use these offsets to calculate the error in depth. Fig. 6.11 shows the error as a function of depth for a stereo camera setup in which the corner offset has been drawn from a Gaussian distribution with variance 0.04.

■ 6.4 Conclusion

In this chapter we have shown how to turn corners into cameras, exploiting a common, but overlooked, visual signal. The vertical edge of a corner's wall selectively blocks light to let the ground nearby display an angular integral of light from around the corner. The resulting penumbras from people and objects are invisible to the eye – typical contrasts are 0.1% above background – but are easy to measure using consumer-grade cameras. We produce 1-D videos of activity around the corner, measured indoors, outdoors, in both sunlight and shade, from brick, tile, wood, and asphalt floors. The resulting 1-D videos reveal the number of people moving around the corner, their angular sizes and speeds, and a temporal summary of activity. Open doorways, with two vertical edges, offer stereo views inside a room, viewable even away from the doorway. Since nearly every corner now offers a 1-D view around the corner, this opens potential applications for automotive pedestrian safety, search and rescue, and public safety. This ever-present, but previously unnoticed, 0.1% signal may invite other novel camera measurement methods.

Appendix A

Static VLBI Imaging Derivations

A.1 Bayesian Inference under Gaussian Distributions

Let,

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{F}\boldsymbol{x}, \boldsymbol{\Sigma}) \tag{A.1}$$

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{A.2}$$

Then,

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) && \text{(A.3)}\\
&= \mathcal{N}_{\boldsymbol{y}}(\boldsymbol{F}\boldsymbol{x}, \Sigma)\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && \text{(A.4)}\\
&\qquad\qquad \text{Using Lemma } D.0.4\\
&= \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{F}^{-1}\boldsymbol{y}, (\boldsymbol{F}^T\Sigma^{-1}\boldsymbol{F})^{-1})\mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) && \text{(A.5)}\\
&\qquad\qquad \text{Using Lemma } D.0.2\\
&= \mathcal{N}_{\boldsymbol{x}}(\hat{\boldsymbol{x}}, \boldsymbol{C}) && \text{(A.6)}
\end{aligned}
$$

for

$$\hat{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{F}^T(\boldsymbol{\Sigma} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}) \tag{A.7}$$

$$\boldsymbol{C} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{F}^T(\boldsymbol{\Sigma} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{\Lambda}. \tag{A.8}$$

A.2 Interstellar Scattering

In) [19] the interstellar scattering kernel is estimated to be a Gaussian with a covariance of

$$\Sigma_{\text{scattering}} = \begin{bmatrix} 2.00614 \times 10^{-21} & 3.21018 \times 10^{-22} \\ 3.21018 \times 10^{-22} & 5.64104 \times 10^{-22} \end{bmatrix} \text{radians.} \tag{A.9}$$

As the Fourier transform of a Gaussian kernel,

$$\frac{1}{2\pi} |\Sigma_{\text{scattering}}|^{-1/2} \exp\left(-\frac{1}{2} [\alpha \quad \delta] \Sigma^{-1}_{\text{scattering}} \begin{bmatrix} \alpha \\ \delta \end{bmatrix} \right), \tag{A.10}$$

is also Gaussian, we can modify the forward model of a visibility by multiplying each $\mathbf{f}(u, v)$ by the scalar,

$$\mathcal{K}(u, v) = \frac{1}{2\pi} |\Sigma_{\text{scattering}}|^{-1/2} \exp\left[-2\pi^2 [u \quad v] \Sigma_{\text{scattering}} \begin{bmatrix} u \\ v \end{bmatrix} \right]. \tag{A.11}$$

In particular, if $\boldsymbol{y}$ is composed of visibilities obtained by observing Sgr A*, then if $\boldsymbol{y}_{2k} = \Re\left[\Gamma(u_k, v_k)\right]$ and $\boldsymbol{y}_{2k+1} = \Im\left[\Gamma(u_k, v_k)\right]$,

$$\begin{aligned} g_{2k}(\hat{I}(\boldsymbol{x})) &= \mathcal{K}(u_k, v_k) \Re[\mathbf{f}(u_k, v_k)]^T \boldsymbol{x} & \text{(A.12)} \\ g_{2k+1}(\hat{I}(\boldsymbol{x})) &= \mathcal{K}(u_k, v_k) \Im[\mathbf{f}(u_k, v_k)]^T \boldsymbol{x}. & \text{(A.13)} \end{aligned}$$

If we do not include the interstellar scatting kernel into our forward model, then we could try deconvolve its effects out in the end. However, this would require the result of the first stage of image reconstruction to produce a blurry image. This image would not necessarily score well under our proposed metric for a "good" looking image. Therefore, sequentially performing these operations often results in artifacts that do not appear if you incorporate the kernel into the forward model. Refer to our paper on scattering mitigation to see further analysis of this trade-off [44].

■ A.3 Approximate Gaussian Noise on the Bispectrum

In this section we present how we approximate the noise on each bispectrum term in $\boldsymbol{y}$ when systematic gain error is negligible.

As discussed in Section 5.2.2, a measured visibility, Γ, is corrupted by thermal and atmospheric phase error in the form

$$\Gamma_{jk} = \hat{\Gamma}_{jk} \exp\left[i(\phi_j - \phi_k)\right] + \epsilon_{jk}, \tag{A.14}$$

for uniformly distributed phase errors, ϕ_j and ϕ_k, and Gaussian distributed thermal noise, ϵ_{jk} on an ideal visibility $\hat{\Gamma}_{jk}$ [29]. Although the phase error due to atmospheric inhomogeneity cancels when using the bispectrum, residual error exists due to thermal noise [124]. Since the bispectrum is the product of three visibilities, its noise distribution is not Gaussian; nonetheless, it can be approximated using a Gaussian which characterizes its first-order noise terms. We determine how to approximate the bispectrum's noise by first expanding the noise terms in the triple product:

$$\begin{aligned}\Gamma_{12}\Gamma_{23}\Gamma_{31} &= \left(\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_2)\right] + \epsilon_{12}\right)\left(\hat{\Gamma}_{23} \exp\left[i(\phi_2 - \phi_3)\right] + \epsilon_{23}\right)\left(\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_1)\right] + \epsilon_{31}\right)\\ &= \hat{\Gamma}_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} + \epsilon_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_2 - \phi_1)\right] + \epsilon_{23}\hat{\Gamma}_{12}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_2)\right] \ldots\\ &+ \epsilon_{31}\hat{\Gamma}_{23}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_3)\right] + \epsilon_{12}\epsilon_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_1)\right] \ldots\\ &+ \epsilon_{23}\epsilon_{31}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_2)\right] + \epsilon_{12}\epsilon_{31}\hat{\Gamma}_{23} \exp\left[i(\phi_2 - \phi_3)\right] + \epsilon_{12}\epsilon_{23}\epsilon_{31}\end{aligned} \tag{A.15}$$

As expected, the resulting expression contains a term, $\hat{\Gamma}_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31}$, which no longer is affected by the introduced phase errors. The first order noise terms,

$$\epsilon_{12}\hat{\Gamma}_{23}\hat{\Gamma}_{31} \exp\left[i(\phi_2 - \phi_1)\right] + \epsilon_{23}\hat{\Gamma}_{12}\hat{\Gamma}_{31} \exp\left[i(\phi_3 - \phi_2)\right] + \epsilon_{31}\hat{\Gamma}_{23}\hat{\Gamma}_{12} \exp\left[i(\phi_1 - \phi_3)\right], \tag{A.16}$$

contribute Gaussian noise to the bispectrum. However, they are each scaled by a value related to the ideal visibilities. Although we are unable to measure these ideal values, we approximate them by using the corresponding measured visibilities. As we expect noise on the bispectrum to be isotropic, we approximate noise on the real and imaginary component of $\Gamma_{12}\Gamma_{23}\Gamma_{31}$ as Gaussian with variance

$$\sigma_{12}^2|\Gamma_{23}\Gamma_{31}|^2 + \sigma_{23}^2|\Gamma_{12}\Gamma_{31}|^2 + \sigma_{31}^2|\Gamma_{12}\Gamma_{23}|^2. \tag{A.17}$$

By comparing our approximation to distributions obtained through sampling we have seen that it accurately models the true noise for SNR values greater than 1. As we expect values of SNR greater than 1 for imaging, a Gaussian noise model is a reasonable approximation for the bispectrum.

To further account for the independence the data in $\boldsymbol{y}$, we scale the k-th term of $\chi(\boldsymbol{x}, \boldsymbol{y})$ in Equation 3.15 by $\alpha_k = \frac{(P_k-1)(P_k-2)}{2\binom{P_k}{3}} = \frac{3}{P_k}$ for P_k telescopes observing at the time corresponding to the bispectrum or closure phase value contained in $\boldsymbol{y}[k]$.

A.4 Data Product Derivatives

In this section A^R and A^I indicate the real and imaginary portions of a complex-valued matrix A, respectively.

A.4.1 Bispectrum

$$\begin{aligned}
A_{i_{1,2}}XA_{i_{2,3}}XA_{i_{1,3}}X &= (A^R_{i_{1,2}} + iA^I_{i_{1,2}})X(A^R_{i_{2,3}} + iA^I_{i_{2,3}})X(A^R_{i_{1,3}} + iA^I_{i_{1,3}})X \\
&= (A^R_{i_{1,2}}X + iA^I_{i_{1,2}}X)(A^R_{i_{2,3}}X + iA^I_{i_{2,3}}X)(A^R_{i_{1,3}}X + iA^I_{i_{1,3}}X) \\
&= A^R_{i_{1,2}}XA^R_{i_{2,3}}XA^R_{i_{1,3}}X - A^R_{i_{1,2}}XA^I_{i_{2,3}}XA^I_{i_{1,3}}X - A^I_{i_{1,2}}XA^R_{i_{2,3}}XA^I_{i_{1,3}}X - A^I_{i_{1,2}}XA^I_{i_{2,3}}XA^R_{i_{1,3}}X + \ldots \\
&i\left(A^R_{i_{1,2}}XA^R_{i_{2,3}}XA^I_{i_{1,3}}X + A^R_{i_{1,2}}XA^I_{i_{2,3}}XA^R_{i_{1,3}}X + A^I_{i_{1,2}}XA^R_{i_{2,3}}XA^R_{i_{1,3}}X - A^I_{i_{1,2}}XA^I_{i_{2,3}}XA^I_{i_{1,3}}X\right) \\
&= \xi_i^R(X) + i\xi_i^I(X)
\end{aligned} \tag{A.18}$$

We must find the derivative of ξ_i^R with respect to X

$$\begin{aligned}
\frac{d}{dX}\xi_i^R &= \frac{d}{dX}\left(A^R_{i_{1,2}}XA^R_{i_{2,3}}XA^R_{i_{1,3}}X - A^R_{i_{1,2}}XA^I_{i_{2,3}}XA^I_{i_{1,3}}X - A^I_{i_{1,2}}XA^R_{i_{2,3}}XA^I_{i_{1,3}}X - A^I_{i_{1,2}}XA^I_{i_{2,3}}XA^R_{i_{1,3}}X\right) \\
&= \frac{d}{dX}(A^R_{i_{1,2}}XA^R_{i_{2,3}}XA^R_{i_{1,3}}X) - \frac{d}{dX}(A^R_{i_{1,2}}XA^I_{i_{2,3}}XA^I_{i_{1,3}}X) - \frac{d}{dX}(A^I_{i_{1,2}}XA^R_{i_{2,3}}XA^I_{i_{1,3}}X) - \frac{d}{dX}(A^I_{i_{1,2}}XA^I_{i_{2,3}}XA^R_{i_{1,3}}X) \\
&= A^R_{i_{1,2}}A^R_{i_{2,3}}XA^R_{i_{1,3}}X + A^R_{i_{2,3}}A^R_{i_{1,2}}XA^R_{i_{1,3}}X + A^R_{i_{1,3}}A^R_{i_{2,3}}XA^R_{i_{1,2}}X\ldots \\
&- (A^R_{i_{1,2}}A^I_{i_{2,3}}XA^I_{i_{1,3}}X + A^I_{i_{2,3}}A^R_{i_{1,2}}XA^I_{i_{1,3}}X + A^I_{i_{1,3}}A^I_{i_{2,3}}XA^R_{i_{1,2}}X)\ldots \\
&- (A^I_{i_{1,2}}A^R_{i_{2,3}}XA^I_{i_{1,3}}X + A^R_{i_{2,3}}A^I_{i_{1,2}}XA^I_{i_{1,3}}X + A^I_{i_{1,3}}A^R_{i_{2,3}}XA^I_{i_{1,2}}X)\ldots \\
&- (A^I_{i_{1,2}}A^I_{i_{2,3}}XA^R_{i_{1,3}}X + A^I_{i_{2,3}}A^I_{i_{1,2}}XA^R_{i_{1,3}}X + A^R_{i_{1,3}}A^I_{i_{2,3}}XA^I_{i_{1,2}}X)
\end{aligned} \tag{A.19}$$

Similarly, the derivative of ξ_i^I with respect to X is

$$\begin{aligned}\frac{d}{dX}\xi_i^I &= \frac{d}{dX}\left(A^R_{i_{1,2}}XA^R_{i_{2,3}}XA^I_{i_{1,3}}X + A^R_{i_{1,2}}XA^I_{i_{2,3}}XA^R_{i_{1,3}}X + A^I_{i_{1,2}}XA^R_{i_{2,3}}XA^R_{i_{1,3}}X - A^I_{i_{1,2}}XA^I_{i_{2,3}}XA^I_{i_{1,3}}X\right)\\ &= A^R_{i_{1,2}}A^R_{i_{2,3}}XA^I_{i_{1,3}}X + A^R_{i_{2,3}}A^R_{i_{1,2}}XA^I_{i_{1,3}}X + A^I_{i_{1,3}}A^R_{i_{2,3}}XA^R_{i_{1,2}}X\ldots\\ &+ A^R_{i_{1,2}}A^I_{i_{2,3}}XA^R_{i_{1,3}}X + A^I_{i_{2,3}}A^R_{i_{1,2}}XA^R_{i_{1,3}}X + A^R_{i_{1,3}}A^I_{i_{2,3}}XA^R_{i_{1,2}}X\ldots\\ &+ A^I_{i_{1,2}}A^R_{i_{2,3}}XA^R_{i_{1,3}}X + A^R_{i_{2,3}}A^I_{i_{1,2}}XA^R_{i_{1,3}}X + A^R_{i_{1,3}}A^R_{i_{2,3}}XA^I_{i_{1,2}}X\ldots\\ &- (A^I_{i_{1,2}}A^I_{i_{2,3}}XA^I_{i_{1,3}}X + A^I_{i_{2,3}}A^I_{i_{1,2}}XA^I_{i_{1,3}}X + A^I_{i_{1,3}}A^I_{i_{2,3}}XA^I_{i_{1,2}}X)\end{aligned} \tag{A.20}$$

■ A.4.2 Visibility Amplitudes

$$\frac{d}{dX}\sqrt{(A^RX)^2+(A^IX)^2} = \frac{d}{dX}\sqrt{A^RXA^RX + A^IXA^IX} = \frac{(A^{RT}(A^RX) + A^{IT}(A^IX))}{\sqrt{(A^RX)^2+(A^IX)^2}} \tag{A.21}$$

Appendix B

StarWarps Derivations

Given a set of observations and model parameters θ, we would like to estimate the marginal distribution for each of the N latent images. For ease of notation, we define:

$$y_{a:b} = \{y_a, ..., y_b\} \tag{B.1}$$

$$y_{a:b \setminus t} = \{y_a, ..., y_{t-1}, y_{t+1}, ..., y_b\} \tag{B.2}$$

According to Equation 4.4, likelihood of the the full joint model is written as:

$$p(x_1, ..., x_N | y_1, ..., y_N) \propto \prod_{t=1}^{N} \mathcal{N}_{x_t}(\mu, \Lambda) \prod_{t=1}^{N} \mathcal{N}_{y_t}(F_t x_t, R_t) \prod_{t=2}^{N} \mathcal{N}_{x_t}(A x_{t-1}, Q) \tag{B.3}$$

■ B.1 Solving for the $E[x_t]$ and $E[x_t x_t^T]$ Sufficient Statistics

We can compute $p(x_t | y_1, ..., y_N)$ by marginalizing out the other latent images:

$$p(x_t | y_{1:N}) \propto \int_{x_{1:N \setminus t}} p(x_{1:N} | y_{1:N}) \;\; dx_1, ..., dx_{1:N \setminus t} \tag{B.4}$$

$$\propto \int_{x_{1:N \setminus t}} \prod_{t=1}^{N} \mathcal{N}_{x_t}(\mu, \Lambda) \prod_{t=1}^{N} \mathcal{N}_{y_t}(F_t x_t, R_t) \prod_{t=2}^{N} \mathcal{N}_{x_t}(A x_{t-1}, Q) \;\; dx_{1:N \setminus t} \tag{B.5}$$

Using the Elimination Algorithm we can solve for this probability efficiently.

$$p(x_t|y_{1:N}) \propto p(x_t, y_{1:N}) \tag{B.6}$$

$$\propto \phi_t(x_t) m_{t-1\to t}(x_t) m_{t+1\to t}(x_t) \tag{B.7}$$

We define the unary and joint potential functions as follows:

$$\phi_t(x_t) = p(y_t|x_t)p(x_t) \tag{B.8}$$

$$= \mathcal{N}_{y_t}(F_t x_t, R_t)\mathcal{N}_{x_t}(\mu, \Lambda) \tag{B.9}$$

Using Lemma D.0.4

$$= \mathcal{N}_{x_t}(F_t^{-1} y_t, (F^T R_t^{-1} F)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda) \tag{B.10}$$

$$\Psi_{t-1,t}(x_{t-1}, x_t) = p(x_t|x_{t-1}) \tag{B.11}$$

$$= \mathcal{N}_{x_t}(Ax_{t-1}, Q) \tag{B.12}$$

Using these potential functions, the resulting sum-product messages are:

$$m_{1\to 2}(x_2) = \int_{x_1} \phi_1(x_1)\Psi_{1,2}(x_1, x_2)dx_1 \tag{B.13}$$

$$m_{t\to t+1}(x_{t+1}) = \int_{x_t} \phi_t(x_t)\Psi_{t,t+1}(x_t, x_{t+1})m_{t-1\to t}(x_t)dx_t \qquad 2 \le t \le N \tag{B.14}$$

$$= \mathcal{N}_{x_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) = \int_{x_t} \Psi_{t,t+1}(x_t, x_{t+1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})dx_t \tag{B.15}$$

$$m_{N\to N-1}(x_2) = \int_{x_N} \phi_N(x_N)\Psi_{N-1,N}(x_{N-1}, x_N)dx_N \tag{B.16}$$

$$m_{t\to t-1}(x_{t-1}) = \int_{x_t} \phi_t(x_t)\Psi_{t-1,t}(x_{t-1}, x_t)m_{t+1\to t}(x_t)dx_t \qquad 1 \le t \le N-1 \tag{B.17}$$

$$= \mathcal{N}_{x_{t-1}}(\mathbf{z}^{\beta}_{t-1|t}, \mathbf{P}^{\beta}_{t-1|t}) = \int_{x_t} \Psi_{t-1,t}(x_{t-1}, x_t)\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})dx_t \tag{B.18}$$

B.1.1 Forward Messages

$$m_{t\to t+1}(x_{t+1}) = \int_{x_t} \Psi_{t,t+1}(x_t, x_{t+1})\phi_t(x_t)m_{t-1\to t}(x_t)dx_t \tag{B.19}$$

$$= \int_{x_t} N_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)m_{t-1\to t}(x_t)dx_t \tag{B.20}$$

$$= \int_{x_t} \mathcal{N}_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})dx_t \tag{B.21}$$

Using Lemma D.0.1: $\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1}) \propto \mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})$

where $\mathbf{z}^{\alpha*}_{1|0} = \mu$, $\mathbf{P}^{\alpha*}_{1|0} = \Lambda$

$$\propto \int_{x_t} \mathcal{N}_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})dx_t \tag{B.22}$$

Using Lemma D.0.2: $\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t}) \propto \mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})$

$$\propto \int_{x_t} \mathcal{N}_{x_t}(A^{-1}x_{t+1}, (A^T Q^{-1} A)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})dx_t \tag{B.23}$$

$$= \mathcal{N}_{A^{-1}x_{t+1}}(\mathbf{z}^{\alpha}_{t|t}, (A^T Q^{-1} A)^{-1} + \mathbf{P}^{\alpha}_{t|t}) \int_{x_t} \mathcal{N}_{x_t}(.,.)dx_t \tag{B.24}$$

$$= \mathcal{N}_{A^{-1}x_{t+1}}(\mathbf{z}^{\alpha}_{t|t}, (A^T Q^{-1} A)^{-1} + \mathbf{P}^{\alpha}_{t|t}) \tag{B.25}$$

Using Lemma D.0.6

$$= \mathcal{N}_{x_{t+1}}(A\mathbf{z}^{\alpha}_{t|t}, Q + A\mathbf{P}^{\alpha}_{t|t}A^T) \tag{B.26}$$

$$= \mathcal{N}_{x_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{B.27}$$

B.1.2 Backward Messages

$$m_{t\to t-1}(x_{t-1}) = \int_{x_t} \Psi_{t-1,t}(x_{t-1}, x_t)\phi_t(x_t)m_{t+1\to t}(x_t)dx_t \tag{B.28}$$

$$= \int_{x_t} \mathcal{N}_{x_t}(Ax_{t-1}, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)m_{t+1\to t}(x_{t-1})dx_t \tag{B.29}$$

$$= \int_{x_t} \mathcal{N}_{x_t}(Ax_{t-1}, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t+1}, \mathbf{P}^{\beta}_{t|t+1})dx_t \tag{B.30}$$

Using Lemma D.0.3: $\mathcal{N}_{x_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1}) \propto \mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t+1}, \mathbf{P}^{\beta}_{t|t+1})$

where $\mathbf{z}^{\beta*}_{N|N+1} = \mu$, $\mathbf{P}^{\beta*}_{N|N+1} = \Lambda$

$$\propto \int_{x_t} \mathcal{N}_{x_t}(Ax_{t-1}, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1})dx_t \tag{B.31}$$

Using Lemma D.0.2: $\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t}) \propto \mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\beta*}_{t|t+1}, \mathbf{P}^{\beta*}_{t|t+1})$

$$\propto \int_{x_t} \mathcal{N}_{x_t}(Ax_{t-1}, Q)\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})dx_t \tag{B.32}$$

$$= \int_{x_t} \mathcal{N}_{Ax_{t-1}}(\mathbf{z}^{\beta}_{t|t}, Q + \mathbf{P}^{\beta}_{t|t})\mathcal{N}_{x_t}(.,.)dx_t \tag{B.33}$$

$$= C\exp\left[\frac{-1}{2}(Ax_{t-1} - \mathbf{z}^{\beta}_{t|t})^T(Q + \mathbf{P}^{\beta}_{t|t})^{-1}(Ax_{t-1} - \mathbf{z}^{\beta}_{t|t})\right]\int_{x_t} \mathcal{N}_{x_t}(.,.)dx_t \tag{B.34}$$

$$= C\exp\left[\frac{-1}{2}(x_{t-1} - A^{-1}\mathbf{z}^{\beta}_{t|t})^T A^T(Q + \mathbf{P}^{\beta}_{t|t})^{-1}A(x_{t+1} - A^{-1}\mathbf{z}^{\beta}_{t|t})\right] \tag{B.35}$$

$$= \mathcal{N}_{x_{t-1}}(A^{-1}\mathbf{z}^{\beta}_{t|t}, (A^T(Q + \mathbf{P}^{\beta}_{t|t})^{-1}A)^{-1}) \tag{B.36}$$

$$= \mathcal{N}_{x_{t-1}}(\mathbf{z}^{\beta}_{t-1|t}, \mathbf{P}^{\beta}_{t-1|t}) \tag{B.37}$$

B.1.3 Putting it Together

$$p(x_t|y_{1:N}) \propto p(x_t, y_{1:N}) \tag{B.38}$$

$$\propto m_{t-1\to t}(x_t)m_{t+1\to t}(x_t)\phi_t(x_t) \tag{B.39}$$

$$= m_{t-1\to t}(x_t)m_{t+1\to t}(x_t)\mathcal{N}_{y_t}(F_t x_t, R_t)\mathcal{N}_{x_t}(\mu, \Lambda) \tag{B.40}$$

$$= m_{t-1\to t}(x_t)m_{t+1\to t}(x_t)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F^T R_t^{-1} F)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda) \tag{B.41}$$

$$= m_{t-1\to t}(x_t)\left[\mathcal{N}_{x_t}(F_t^{-1}y_t, (F^T R_t^{-1} F)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)m_{t+1\to t}(x_t)\right] \tag{B.42}$$

$$\propto \mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t})) \tag{B.43}$$

$$\propto \mathcal{N}_{x_t}(\mathbf{z}, \mathbf{P}) \tag{B.44}$$

where using Lemma D.0.1

$$\mathbf{z_t} = \mathbf{P}^{\beta}_{t|t}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\alpha}_{t|t-1} + \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{z}^{\beta}_{t|t} \tag{B.45}$$

$$\mathbf{P_t} = \mathbf{P}^{\alpha}_{t|t-1}(\mathbf{P}^{\alpha}_{t|t-1} + \mathbf{P}^{\beta}_{t|t})^{-1}\mathbf{P}^{\beta}_{t|t} \tag{B.46}$$

Note that for $t = 1$ $p(x_t|y_{1:N}) \propto \mathcal{N}_{x_t}(\mathbf{z}^{\beta}_{t|t}, \mathbf{P}^{\beta}_{t|t}))$. Note that $E[x_t] = \mathbf{z}$. The statistic $E[x_t x_t^T]$ can be also solved using these terms:

$$E_{x_t|y_{1:N},\theta^{(i)}}\left[x_t x_t^T|y_{1:N}\right] = \int_{x_{1:N}} x_t x_t^T p(x_{1:N}|y_{1:N})dx_{1:N} \tag{B.47}$$

$$= \int_{x_t} x_t x_t^T \int_{x_{1:N\setminus t}} p(x_{1:N}|y_{1:N})dx_{1:N\setminus t}dx_t \tag{B.48}$$

$$= \int_{x_t} x_t x_t^T p(x_t|y_{1:N})dx_t \tag{B.49}$$

$$= \int_{x_t} x_t x_t^T \mathcal{N}_{x_t}(z_t, P_t)dx_t \tag{B.50}$$

$$= z_t z_t^T + P_t \tag{B.51}$$

■ B.2 Solving for the $E[x_{t-1}x_t^T]$ Sufficient Statistic

$$p(x_t, x_{t-1}|y_{1:N}) \propto \int_{x_{1:N\setminus t,t-1}} p(x_{1:N}|y_{1:N}) \ \ dx_{1:N\setminus t,t-1} \tag{B.52}$$

$$\propto \int_{x_{1:N\setminus t,t-1}} \prod_{t=1}^{N} \mathcal{N}_{x_t}(\mu, \Lambda) \prod_{t=1}^{N} \mathcal{N}_{y_t}(F_t x_t, R_t) \prod_{t=2}^{N} \mathcal{N}_{x_t}(Ax_{t-1}, Q) \ \ dx_{1:N\setminus t,t-1} \tag{B.53}$$

Using the Elimination Algorithm we can solve for this probability efficiently.

$$
\begin{align}
p(x_t, x_{t-1}|y_{1:N}) \propto\, & p(x_t, x_{t-1}, y_{1:N}) \tag{B.54}\\
&\propto m_{t-2\to t-1}\phi_{t-1}(x_{t-1})\Psi_{t-1,t}(x_{t-1}, x_t)\phi_t(x_t)(x_{t-1})m_{t+1\to t}(x_t) \tag{B.55}\\
&= m_{t-2\to t-1}\mathcal{N}_{x_{t-1}}(F_{t-1}^{-1}y_{t-1}, (F_{t-1}^T R_{t-1}^{-1} F_{t-1})^{-1})\mathcal{N}_{x_{t-1}}(\mu, \Lambda)\mathcal{N}_{x_t}(Ax_{t-1}, Q) \notag\\
&\qquad\qquad \mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)(x_{t-1})m_{t+1\to t}(x_t) \tag{B.56}\\
&= \mathcal{N}_{x_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{x_t}(Ax_{t-1}, Q)\mathcal{N}_{x_t}(\mathbf{z}^\beta_{t|t}, \mathbf{P}^\beta_{t|t}) \tag{B.57}\\
&= \mathcal{N}_{x_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{Ax_{t-1}}(\mathbf{z}^\beta_{t|t}, Q + \mathbf{P}^\beta_{t|t}) \tag{B.58}\\
&\quad \mathcal{N}_{x_t}(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(Q+\mathbf{P}^\beta_{t|t})^{-1}Ax_{t-1}, Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) \tag{B.59}\\
&= \mathcal{N}_{x_{t-1}}(\mathbf{z}^\alpha_{t-1|t-1}, \mathbf{P}^\alpha_{t-1|t-1})\mathcal{N}_{x_{t-1}}(A^{-1}\mathbf{z}^\beta_{t|t}, (A^T(Q+\mathbf{P}^\beta_{t|t})^{-1}A)^{-1}) \tag{B.60}\\
&\quad \mathcal{N}_{x_t}(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(Q+\mathbf{P}^\beta_{t|t})^{-1}Ax_{t-1}, Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) \tag{B.61}\\
&= K_t\mathcal{N}_{x_{t-1}}(m_{t-1}, C_{t-1}) \notag\\
&\quad \mathcal{N}_{x_t}(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} + \mathbf{P}^\beta_{t|t}(Q+\mathbf{P}^\beta_{t|t})^{-1}Ax_{t-1}, Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) \tag{B.62}\\
&= K_t\mathcal{N}_{x_{t-1}}(m_{t-1}, C_{t-1}) \notag\\
&\quad \mathcal{N}_{\mathbf{P}^\beta_{t|t}(Q+\mathbf{P}^\beta_{t|t})^{-1}Ax_{t-1}}(x_t - Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t}, Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}) \tag{B.63}\\
&= K_t\mathcal{N}_{x_{t-1}}(m_{t-1}, C_{t-1}) \notag\\
&\quad \mathcal{N}_{x_{t-1}}(M^{-1}(x_t - Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t}), ((M^T(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t})^{-1}M)^{-1}) \tag{B.64}\\
&= K_t\mathcal{N}_{x_{t-1}}(m_{t-1} + C_{t-1}M^T(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1} \notag\\
&\qquad\qquad (x_t - Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} - Mm_{t-1}), \tag{B.65}\\
&\qquad\qquad C_{t-1} - C_{t-1}M^T(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1}MC_{t-1}) \tag{B.66}\\
&= N_{x_{t-1}}(Dx_t + G, F) \tag{B.67}
\end{align}
$$

$$
\begin{align}
M &= \mathbf{P}^\beta_{t|t}(Q+\mathbf{P}^\beta_{t|t})^{-1}A \tag{B.68}\\
m_{t-1} &= \mathbf{z}^\alpha_{t-1|t-1} + \mathbf{P}^\alpha_{t-1|t-1}A^T(Q+\mathbf{P}^\beta_{t|t} + A\mathbf{P}^\alpha_{t-1|t-1}A^T)^{-1}(\mathbf{z}^\beta_{t|t} - A\mathbf{z}^\alpha_{t-1|t-1}) \tag{B.69}\\
C_{t-1} &= \mathbf{P}^\alpha_{t-1|t-1} - \mathbf{P}^\alpha_{t-1|t-1}A^T(Q+\mathbf{P}^\beta_{t|t} + A\mathbf{P}^\alpha_{t-1|t-1})A^T)^{-1}A\mathbf{P}^\alpha_{t-1|t-1} \tag{B.70}\\
D &= C_{t-1}M^T(Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t} + MC_{t-1}M^T)^{-1} \tag{B.71}\\
G &= m_{t-1} + D(-Q(Q+\mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t} - Mm_{t-1}) \tag{B.72}\\
F &= C_{t-1} - DMC_{t-1} \tag{B.73}
\end{align}
$$

Using the following relations we can then compute $E[x_{t-1}x_t^T]$:

$$E[y(Ax+B)^T] = E[yx^TA^T + yB^T] = E[yx^T]A^T + E[y]B^T \tag{B.74}$$

$$\begin{aligned}
F + E[x_{t-1}]E[Dx_t + G]^T =& E[x_{t-1}(Dx_t + G)^T] = E[x_{t-1}x_t^T]D^T + E[x_{t-1}]G^T && \text{(B.75)}\\
E[x_{t-1}x_t^T]D^T =& (F + E[x_{t-1}]E[Dx_t + G]^T - E[x_{t-1}]G^T) && \text{(B.76)}\\
E[x_{t-1}x_t^T] =& (F + E[x_{t-1}]E[Dx_t + G]^T - E[x_{t-1}]G^T)D^{T-1} && \text{(B.77)}\\
E[x_{t-1}x_t^T] =& (F + E[x_{t-1}]E[x_t]^TD^T + E[x_{t-1}]G^T - E[x_{t-1}]G^T)D^{T-1} && \\
& && \text{(B.78)}\\
E[x_{t-1}x_t^T] =& (F + E[x_{t-1}]E[x_t]^TD^T)D^{T-1} && \text{(B.79)}\\
E[x_{t-1}x_t^T] =& FD^{T-1} + E[x_{t-1}]E[x_t]^T && \text{(B.80)}
\end{aligned}$$

■ B.3 Likelihood

$$\begin{aligned}
p(y_{1:N}) &= \int_{x_{1:N}} p(x_{1:N}, y_{1:N}) \; dx_{1:N} && \text{(B.81)}\\
&= \int_{x_{1:N}} \prod_{t=1}^{N} \mathcal{N}_{x_t}(\mu, \Lambda) \prod_{t=1}^{N} \mathcal{N}_{y_t}(F_t x_t, R_t) \prod_{t=2}^{N} \mathcal{N}_{x_t}(Ax_{t-1}, Q) \; dx_{1:N} && \text{(B.82)}
\end{aligned}$$

Note that this can be solved with message passing. This is very similar to the calculations done for the forward message passing. The only difference is that it also keeps tracks of scaling coefficients.

$$m_{t\to t+1}(x_{t+1}) = \int_{x_t} \Psi_{t,t+1}(x_t, x_{t+1})\phi_t(x_t)m_{t-1\to t}(x_t)dx_t \tag{B.83}$$

$$= \int_{x_t} N_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)m_{t-1\to t}(x_t)dx_t \tag{B.84}$$

$$= \int_{x_t} \mathcal{N}_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda)\ell_t\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})dx_t \tag{B.85}$$

Using Lemma D.0.1 and Matrix Cookbook [105]:

$$\mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{t|t-1}, \Lambda + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1}) = \mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t-1}, \mathbf{P}^{\alpha}_{t|t-1})$$

where $\mathbf{z}^{\alpha*}_{1|0} = \mu$, $\mathbf{P}^{\alpha*}_{1|0} = \Lambda$

$$= \ell_t\mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{t|t-1}, \Lambda + \mathbf{P}^{\alpha}_{t|t-1}) \int_{x_t} \mathcal{N}_{x_{t+1}}(Ax_t, Q)\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})dx_t \tag{B.86}$$

Using Matrix Cookbook [105], Lemma D.0.2, and Lemma D.0.6:

$$\mathcal{N}_{x_t}(F_t^{-1}y_t, (F_t^T R_t^{-1} F_t)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha*}_{t|t-1}, \mathbf{P}^{\alpha*}_{t|t-1})$$

$$= \mathcal{N}_{F_t^{-1}y_t}(\mathbf{z}^{\alpha*}_{t|t-1}, (F_t^T R_t^{-1} F_t)^{-1} + \mathbf{P}^{\alpha*}_{t|t-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})$$

$$= \mathcal{N}_{y_t}(F\mathbf{z}^{\alpha*}_{t|t-1}, R_t + F_t\mathbf{P}^{\alpha*}_{t|t-1}F^T)\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})$$

$$= \ell_t\mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{t|t-1}, \Lambda + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{y_t}(F\mathbf{z}^{\alpha*}_{t|t-1}, R_t + F_t\mathbf{P}^{\alpha}_{t|t-1}F^T) \int_{x_t} \mathcal{N}_{x_t}(A^{-1}x_{t+1}, (A^T Q^{-1} A)^{-1})\mathcal{N}_{x_t}(\mathbf{z}^{\alpha}_{t|t}, \mathbf{P}^{\alpha}_{t|t})dx_t \tag{B.87}$$

$$= \ell_t\mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{t|t-1}, \Lambda + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{y_t}(F\mathbf{z}^{\alpha*}_{t|t-1}, R_t + F_t\mathbf{P}^{\alpha}_{t|t-1}F^T)\mathcal{N}_{x_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{B.88}$$

$$= \ell_{t+1}\mathcal{N}_{x_{t+1}}(\mathbf{z}^{\alpha}_{t+1|t}, \mathbf{P}^{\alpha}_{t+1|t}) \tag{B.89}$$

$$\ell_{t+1} = \ell_t\mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{t|t-1}, \Lambda + \mathbf{P}^{\alpha}_{t|t-1})\mathcal{N}_{y_t}(F_t\mathbf{z}^{\alpha*}_{t|t-1}, R_t + F_t\mathbf{P}^{\alpha*}_{t|t-1}F_t^T) \tag{B.90}$$

Initialization:

$$\ell_0 = 1 \ , \ \mathcal{N}_{\mu}(\mathbf{z}^{\alpha}_{1|0}, \Lambda + \mathbf{P}^{\alpha}_{1|0}) = 1 \tag{B.91}$$

Algorithm 3. Computing Likelihood: $t = 1 \to 2 \to \dots \to N$

B.4 Non-linear Observation Model

Now, let's say that our observation model is no longer linear, but instead is a non-linear function of x.

$$y_t \sim \mathcal{N}(F(x_t), R) \tag{B.92}$$

Using a first order Taylor Series Expansion Approximation, we can approximate $\mathcal{N}_y(F(x), R)$ as

$$\begin{aligned}
\mathcal{N}_y(F(x), R) &\propto \exp\left[\frac{-1}{2}(y - F(x))^T R^{-1}(y - F(x))\right] && \text{(B.93)}\\
&\approx \exp\left[\frac{-1}{2}(y - (F(\tilde{x}) + F'(\tilde{x})(x - \tilde{x})))^T R^{-1}(y - (F(\tilde{x}) + F'(\tilde{x})(x - \tilde{x})))\right]\\
&= \exp\left[\frac{-1}{2}(y - F(\tilde{x}) - F'(\tilde{x})x + F'(\tilde{x})\tilde{x})^T R^{-1}(y - F(\tilde{x}) - F'(\tilde{x})x + F'(\tilde{x})\tilde{x})\right]\\
&= \exp\left[\frac{-1}{2}(-1 \times (F'(\tilde{x})x - (-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y)))^T R^{-1}(-1 \times (F'(\tilde{x})x - (-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y)))\right]\\
&= \exp\left[\frac{-1}{2}(F'(\tilde{x})x - (-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y))^T R^{-1}(F'(\tilde{x})x - (-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y))\right]\\
&= \exp\left[\frac{-1}{2}(x - F'(\tilde{x})^{-1}(-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y))^T F'(\tilde{x})^T R^{-1} F'(\tilde{x})(x - F'(\tilde{x})^{-1}(-F(\tilde{x}) + F'(\tilde{x})\tilde{x} + y))\right]\\
&\propto \mathcal{N}_x(F'(\tilde{x})^{-1}(y + F'(\tilde{x})\tilde{x} - F(\tilde{x})), (F'(\tilde{x})^T R^{-1} F'(\tilde{x}))^{-1}) && \text{(B.94)}
\end{aligned}$$

Updates are very similar to in the linear model. We simply need to replace

$$F_t \rightarrow F_t'(\tilde{x}_t) \tag{B.95}$$

$$y_t \rightarrow y_t + F_t'(\tilde{x}_t)\tilde{x}_t - F_t(\tilde{x}_t) \tag{B.96}$$

where $\tilde{x}_t$ is the current estimate of x_t

B.4.1 Gamma Regularization

Often we would like to pull out regions of flux that are orders of magnitude smaller than the peak value in the image. For instance, we may be more interested in the distribution of flux in the log image than in the linear scale intensity image. To deal with this in this framework, we can perform a change of variables and regularize a non-linear-scale image using a Gaussian distribution. Let, $\breve{x}$ be the linear scale image and x be a non-linear-scale image. Gamma correction is often used to approximate a logarithmic function, so we use it here for that purpose:

$$x = \breve{x}^{(1/\gamma)} \tag{B.97}$$

$$\breve{x} = x^{\gamma} \tag{B.98}$$

Now,

$$f(x) = F\breve{x} = Fx^{\gamma} \tag{B.99}$$

$$F \rightarrow \frac{d}{dx} f(x) = F \; Diag(\gamma \hat{x}^{\gamma-1}) \tag{B.100}$$

$$y_t \rightarrow y_t + F \; Diag(\gamma \hat{x}^{\gamma-1})\hat{x} - F\hat{x}^{\gamma} \tag{B.101}$$

A $\gamma = 2.0$ empirically works well in practice.

B.5 Estimating the Evolution Parameters

Calculate the expected value of $\log \Pi_t P(y_t|x_t;\theta)$ with respect to the distribution of x

$$Q(\theta|\theta^{(i)}) = E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[\log P(x_{1:N}, y_{1:N}; \theta)\right] \tag{B.102}$$

$$\begin{aligned}
&= E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[\frac{-1}{2} \sum_{t=2}^{N} \left[(x_t - A_{\theta^{(i)}} x_{t-1})^T Q^{-1} (x_t - A_{\theta^{(i)}} x_{t-1}) \right] + \mathcal{G}(x_{1:N}, y_{1:N}) \right] \\
&= E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[\frac{-1}{2} \sum_{t=2}^{N} \left[x_t^T Q^{-1} x_t - x_t^T Q^{-1} A_{\theta^{(i)}} x_{t-1} - x_{t-1}^T A_{\theta^{(i)}}^T Q^{-1} x_t + x_{t-1}^T A_{\theta^{(i)}}^T Q^{-1} A_{\theta^{(i)}} x_{t-1} \right] + \mathcal{G}(x_{1:N}, y_{1:N}) \right] \\
&= \frac{-1}{2} \sum_{t=2}^{N} \Big[E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[x_t^T Q^{-1} x_t \right] - E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[x_t^T Q^{-1} A_{\theta^{(i)}} x_{t-1} \right] - E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[x_{t-1}^T A_{\theta^{(i)}}^T Q^{-1} x_t \right] \\
&\qquad + E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[x_{t-1}^T A_{\theta^{(i)}}^T Q^{-1} A_{\theta^{(i)}} x_{t-1} \right] + E_{x_{1:N}|y_{1:N},\theta^{(i)}} \left[\mathcal{G}(x_{1:N}, y_{1:N}) \right] \Big]
\end{aligned} \tag{B.103}$$

Where $\mathcal{G}(x_{1:N}, y_{1:N})$ is a term that incorporates the prior for X as well as the terms to normalize each of the distributions. Next, we must find the A that maximizes $Q(\theta|\theta^{(i)})$. Using the Matrix Cookbook (70) and (88) [105]

$$\begin{aligned}\frac{d}{dA_{\theta^{(i)}}}Q(\theta|\theta^{(i)}) &= \frac{1}{2}\sum_{t=2}^{N}\Big[E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[(x_t^T Q^{-1})^T x_{t-1}^T\right] + E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[(x_{t-1}^T Q^{-1})^T x_t^T\right] \\ &\qquad - E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[2Q^{-1}A_{\theta^{(i)}}x_{t-1}x_{t-1}^T\right]\Big] \\ &= \frac{1}{2}\sum_{t=2}^{N}\Big[Q^{-1}E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[x_t x_{t-1}^T\right] + Q^{-1}E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[x_{t-1}x_t^T\right] - 2Q^{-1}A_{\theta^{(i)}}E_{x_{1:N}|y_{1:N},\theta^{(i)}}\left[x_{t-1}x_{t-1}^T\right]\Big] \\ &= \sum_{t=2}^{N}\left[Q^{-1}\left(\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t}\right) - Q^{-1}A_{\theta^{(i)}}M_{t-1}\right] \qquad \text{(B.104)}\end{aligned}$$

$$A_{\theta^{(i+1)}} = \left[\sum_{t=2}^{N}\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t} - Bz_t^T\right]\left[\sum_{t=2}^{N}M_{t-1}\right]^{-1} \qquad \text{(B.105)}$$

Now Let's say that A is a function of θ and we would like to solve for the best θ. To do this we use the Chain rule. From the Matrix Cookbook [105] we know

$$\frac{dQ}{d\theta_j} = \sum_p\sum_q \frac{dQ}{dA_{p,q}}\frac{dA_{p,q}}{d\theta_j} \qquad \text{(B.106)}$$

$$\frac{d}{d\theta_j^{(i)}}Q(\theta|\theta^{(i)}) = \sum_p\sum_q \frac{d}{dA_{p,q}}Q(\theta|\theta^{(i)})\frac{dA_{p,q}}{d\theta_j^{(i)}} \qquad \text{(B.107)}$$

$$= \sum_p\sum_q\left[\sum_{t=2}^{N}\left[Q^{-1}\left(\frac{1}{2}M_{t,t-1} + \frac{1}{2}M_{t-1,t}\right) - Q^{-1}A_{\theta^{(i)}}M_{t-1} - Q^{-1}Bz_t^T\right]\right]_{p,q}\left[\frac{dA(\theta^{(i)})}{d\theta_j^{(i)}}\right]_{p,q} \qquad \text{(B.108)}$$

Appendix C

Corner Camera Derivations

C.1 Corner Location Errors

One important source of error in the edge camera idea is the *corner location error.* When studying a movie of the observation plane, it's important to know where the corner of the wall is in order to make an accurate reconstruction. Corner location errors occur when the corner of the wall is erroneously chosen to be the wrong place. They introduce systematic error into the scene's reconstruction.

C.1.1 Edge Camera

Exactly how bad are corner location errors? To answer this question, we consider the situation shown in Fig. C.1. Imagine a dark scene with a single bright object. We want to find the angular position of the bright object in the scene. We can do this by measuring θ: the angle of the shadow it casts against the wall. When we find the angle at which the observation plane goes from light to dark, we will know what θ is.

This story is simple in the case when there is no corner location error. But what about the case where there is such an error?

Fig. C.2 shows this scenario. We can "sweep" the angle ϕ across the observation plane, and at the point where ϕ is midway between dark and light, we can presume that that will be the most likely inferred angular position of the object from the reconstructed space-time image. When there is no corner location error, we will naturally get $\theta = \phi$, but when there is a corner location error, ϕ will not equal θ, but will depend on θ and other parameters.

Fig. C.3 plots intensity against the sweeping angle ϕ, both with and without a corner location error. Note that in the case where there is a corner location error, the maximum intensity value no longer takes on a maximum value of 1, but a value below 1. In the analysis that follows, we will call that value $l_{\max}$, and we will use $l_{\max}/2$ as

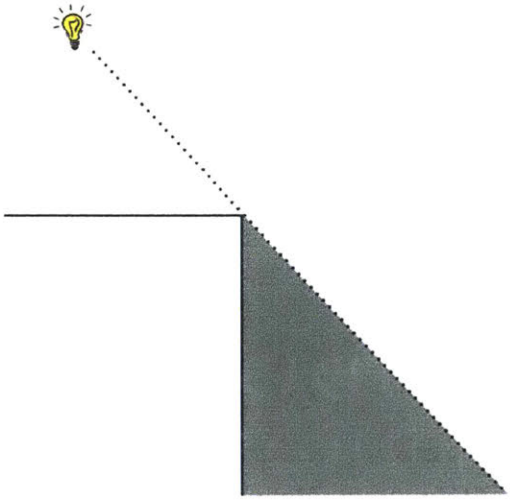


Figure C.1. This figure shows the setup for the toy problem of interest. The scene consists of a single bright object, whose angular position θ we want to learn.

the "transition point" between light and dark. In other words, we will choose the ϕ that gives an intensity of $l_{\max}/2$ as our estimate for θ.

Fig. C.4 is a detailed illustration of the situation, showing the names for the variables that we'll use in our analysis. As the figure shows, we are presuming a corner location error of (d_x, d_y) and a observation plane radius of r. We want to find what our estimate θ, ϕ, will be as a function of θ and in terms of d_x, d_y, and r.

Using Fig. C.4 as a reference, we can make the following observations:

$$\beta = \tan^{-1}\left(\frac{d_x}{d_y}\right)$$

$$l_{\max} = r - d_y \tan(\theta) + d_x$$

$$f = r - \frac{l_{\max}}{2}$$

$$\alpha = \sin^{-1}\left(\frac{\sqrt{d_x^2 + d_y^2}\,\sin(\theta - \beta)}{f}\right)$$

$$\gamma = \pi - \alpha - \theta + \beta$$

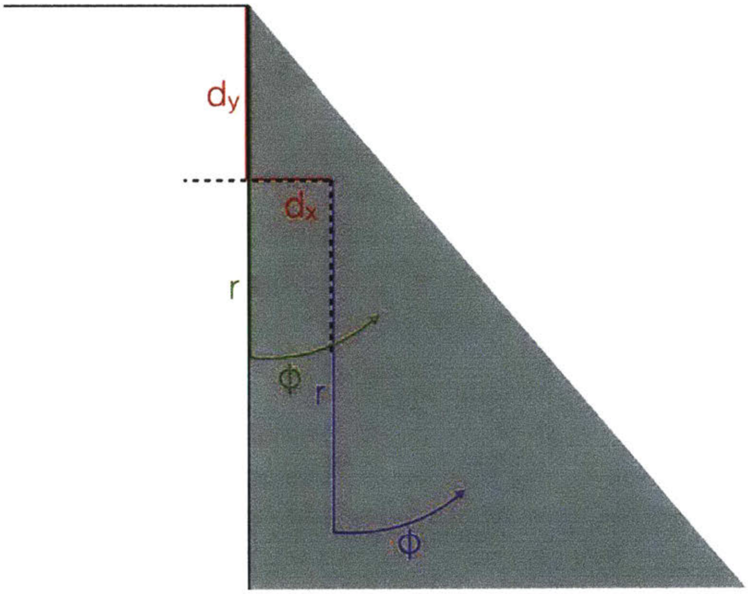


Figure C.2. This figure shows the impact of a corner location error. In the error-free case, we would sweep ϕ across the observation plane (shown in green) hinging around the corner (the solid black line). But if made a corner location error, we would instead try to sweep ϕ across the observation plane erroneously (shown in blue) hinging around the false corner (shown with a dotted line).

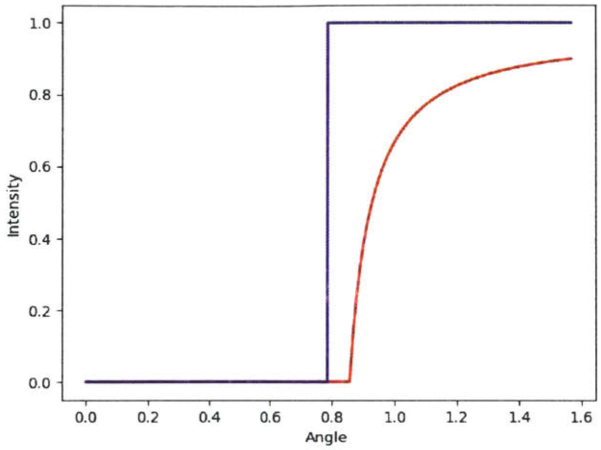


Figure C.3. This plot shows how the observed intensity values vary with ϕ, in the case of correct corner location (in blue) and a corner location error ($(d_x, d_y) = (0.1, 0.2)$, in red). Note that in the case of a corner location error, the maximum value of the intensity does not reach 1. Note also that d_x and d_y are as a fraction of the radius of the observation plane, r, which here is taken to be 1.

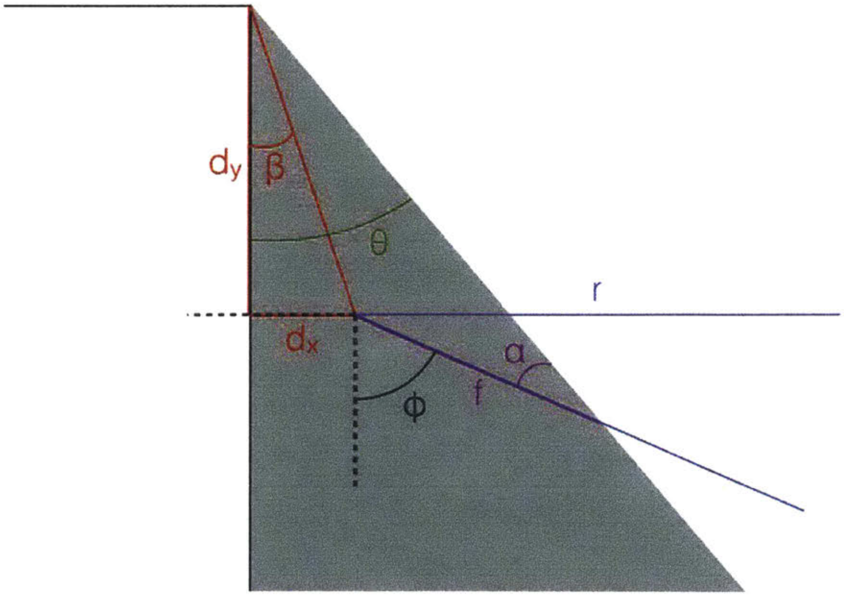


Figure C.4. This plot is intended as a reference for the meanings of each of the variables used in the calculations of ϕ as a function of θ.

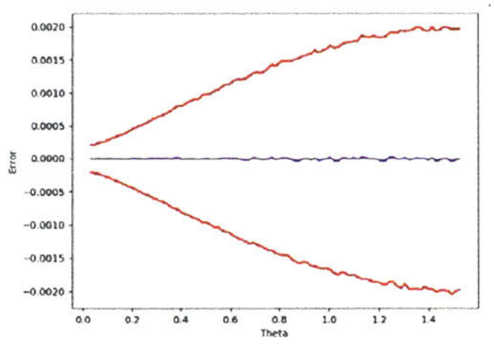


Figure C.5. This plot shows the empirical mean (in blue) plus or minus one standard deviation (in red) of the error as a function of θ. Here, $\sigma_x = 10^{-4}$ and $\sigma_y = 10^{-3}$.

$$\phi = \pi - \gamma + \beta$$

This is how ϕ is expressed in terms of the parameters of the problem (θ, d_x, d_y, r).

What sort of error does this introduce? In order to study this question, we assumed that d_x and d_y were normally distributed with means of 0 and small (relative to r^2) variances σ_x^2 and σ_y^2. We generated many sample $(\theta,\ \phi)$ pairs for each θ between 0 and $\pi/2$. We then measured the empirical means and variances of these pairs. Fig C.5 shows a few of our results.

Here were a few of our empirical findings:

1. The mean error was always 0 for all values of θ, σ_x and σ_y.
2. When $\sigma_x = \sigma_y$, the standard deviation of the error σ_ϵ was $2\sigma_x$ for all values of θ.
3. When $\sigma_x \neq \sigma_y$, the standard deviation of the error σ_ϵ varied between $2\sigma_x$ (for $\theta = 0$) and $2\sigma_y$ (for $\theta = \pi/2$).

■ C.1.2 Stereo Camera

Another situation in which it makes sense to study corner location errors is the case where there is a doorway just before the hidden scene, in which case we can use stereo vision to locate a moving object in two dimensions. We would like to know: what effect do corner location errors have on depth estimates, which are generally quite sensitive to noise? To be more precise, suppose that we call the axis along which the doorway lies the "x-axis," and suppose we call the perpendicular axis (of depth into the room) the

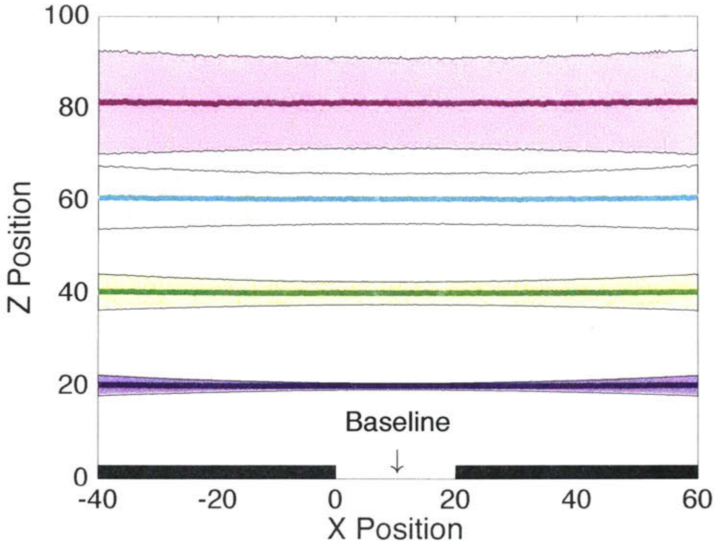


Figure C.6. The empirical means plus or minus one standard deviation of the estimated P_z as a function of its x-coordinate, assuming true P_z of 20, 40, 60, and 80. Here, the two corner location errors at each of the boundaries of the doorway are independent and subject to $\sigma^2_{\Delta x} = \sigma^2_{\Delta z} = 0.04$. We sample from a set of 1000 corner errors to approximate the mean and standard deviations empirically.

"z-axis." Then, how much noise in the z dimension will a corner location error cause?

To give an approximate sense of how much error results in the recovered z position, we show the mean +/- one standard deviation in the z dimension as a function of the true x-position of the object in Fig. C.6

Note that the empirical means are centered at the true depths of the objects. This does *not* mean that any single corner location error won't cause the depth of the reconstruction to be off systematically; it only means that on average, corner location errors that are normally distributed around the corners in question will push the reconstructed depths away as much as they pull them closer.

To see this systematic bias on its own, we can also study how a single corner error introduces systematic error in our reconstructions—after all, for a single experiment, we are likely to make a single corner error, and the resulting error in the depth calculations will extend across many x-coordinates as the subject of the experiment walks back and forth in the hidden scene. Figs. C.7 and C.8 show the systematic bias for two distinct *specific* corner location errors.

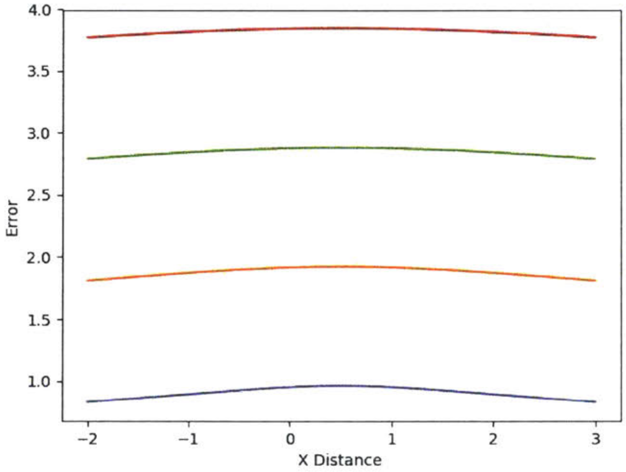


Figure C.7. The reconstructed depths of objects at depths 1, 2, 3, and 4, given a corner error of $\Delta y_1 = \Delta y_2 = 0.02$.

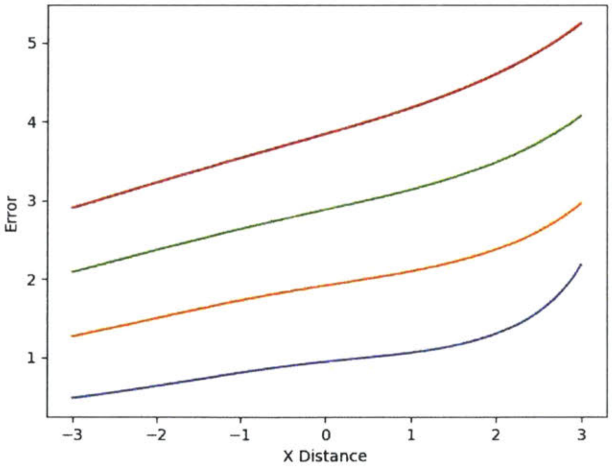


Figure C.8. The reconstructed depths of objects at depths 1, 2, 3, and 4, given a corner error of $\Delta y_1 = -0.02$, $\Delta y_2 = 0.02$. Note that because of the different corner errors for each corner, there is the possibility of asymmetric behavior on either side of the doorway.

■ C.2 Temporal Smoothing

In addition to spatial smoothness we can also reduce noise temporally by imposing smoothness on our MAP estimate. $\hat{\mathbf{x}}^{(t)}$, or averaging adjacent frames in time. This helps to reduce noise, at the cost of some temporal blurring. However, to emphasis the coherence among results, we did not previously impose this additional constraint. Each 1-D image, $\mathbf{x}$, that we showed in the main paper was independently computed. In Figure C.9 we show some results obtained with temporal smoothing. We either computed the 1-D video on averaged video frames, or applied Kalman Filtering/Smoothing to temporally smooth the result. Note that the Kalman Filtering/Smoothing causes the noise to blur out into vertical lines. Frame averaging results appear noiser only because there are less 1D reconstructed frames being shown.

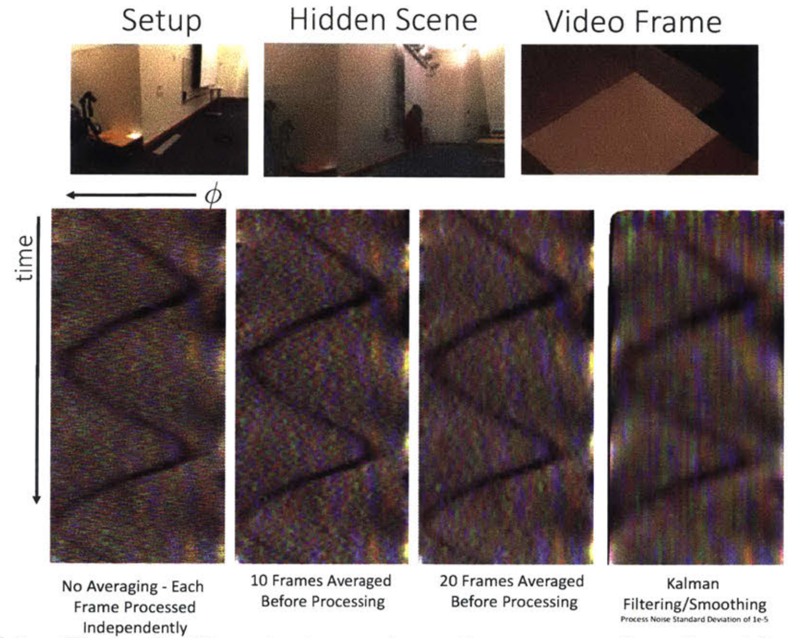


Figure C.9. The result of imposing temporal smoothness or averaging adjacent frames in time to help in reducing noise.

■ C.2.1 Kalman Filtering/Smoothing

We model our scene as a linear dynamic system.

$$x_t = Fx_{t-1} + w_k \tag{C.1}$$

$$y_t = Ax_t + n_t \tag{C.2}$$

$$w_k \sim \mathcal{N}(0, R) \tag{C.3}$$

$$n_t \sim \mathcal{N}(0, \lambda \mathbb{1}) \tag{C.4}$$

The matrix F can be set as the identity matrix so changes over time are penalized. The marginals $p(x_t|y_1, ..., y_n)$ of this HMM can then be solved using the forward backward algorithm - or equivalently Kalman Filtering and Smoothing.

Appendix D

General Lemmas

D.0.1 Lemma 1

$$\mathcal{N}_{x_t}(\mathbf{z}^*_{t|t-1}, \mathbf{P}^*_{t|t-1}) = \mathcal{N}_{x_t}(\mu, \Lambda)\mathcal{N}_{x_t}(\mathbf{z}_{t|t-1}, \mathbf{P}_{t|t-1}) \propto \mathcal{N}_{x_t}(m_t, C_t) \tag{D.1}$$

where from the Matrix Cookbook (8.1.8) and (3.2.5) [105] we know that

$$\begin{aligned}
C_t &= (\Lambda^{-1} + \mathbf{P}^{-1}_{t|t-1})^{-1} && \text{(D.2)}\\
&= \Lambda(\Lambda + \mathbf{P}_{t|t-1})^{-1}\mathbf{P}_{t|t-1} && \text{(D.3)}\\
m_t &= (\Lambda^{-1} + \mathbf{P}^{-1}_{t|t-1})^{-1}(\Lambda^{-1}\mu + \mathbf{P}^{-1}_{t|t-1}\mathbf{z}_{t|t-1}) && \text{(D.4)}\\
&= \Lambda(\Lambda + \mathbf{P}_{t|t-1})^{-1}\mathbf{z}_{t|t-1} + \mathbf{P}_{t|t-1}(\Lambda + \mathbf{P}_{t|t-1})^{-1}\mu && \text{(D.5)}\\
& && \text{(D.6)}
\end{aligned}$$

D.0.2 Lemma 2

$$\mathcal{N}_x(\mathbf{A}^{-1}y, (\mathbf{A}^T\Sigma^{-1}\mathbf{A})^{-1})\mathcal{N}_x(\mu, Q) \propto \mathcal{N}_x(m, C) \tag{D.7}$$

where from the Matrix Cookbook (8.1.8) [105] and the Woodbury matrix identity we know that

$$C_t = (\mathbf{A}^T\Sigma^{-1}\mathbf{A} + Q^{-1})^{-1} \tag{D.8}$$

$$= Q - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}Q \tag{D.9}$$

$$m_t = C_t(\mathbf{A}^T\Sigma^{-1}\mathbf{A}\mathbf{A}^{-1}y + Q^{-1}\mu) \tag{D.10}$$

$$= (Q - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}Q)(\mathbf{A}^T\Sigma^{-1}y + Q^{-1}\mu) \tag{D.11}$$

$$= Q\mathbf{A}^T\Sigma^{-1}y + QQ^{-1}\mu - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}Q\mathbf{A}^T\Sigma^{-1}y - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}QQ^{-1}\mu) \tag{D.12}$$

$$= \mu + Q\mathbf{A}^T\Sigma^{-1}y - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}Q\mathbf{A}^T\Sigma^{-1}y - Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}\mathbf{A}\mu) \tag{D.13}$$

$$= \mu + Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}((\Sigma + \mathbf{A}Q\mathbf{A}^T)\Sigma^{-1}y - \mathbf{A}Q\mathbf{A}^T\Sigma^{-1}y - \mathbf{A}\mu) \tag{D.14}$$

$$= \mu + Q\mathbf{A}^T(\Sigma + \mathbf{A}Q\mathbf{A}^T)^{-1}(y - \mathbf{A}\mu) \tag{D.15}$$

■ D.0.3 Lemma 3

$$\mathcal{N}_{x_t}(\mathbf{z}^*_{t|t+1}, \mathbf{P}^*_{t|t+1}) = \mathcal{N}_{x_t}(\mathbf{z}_{t|t+1}, \mathbf{P}_{t|t+1})\mathcal{N}_{x_t}(\mu, \Lambda) \tag{D.16}$$

$$= \mathcal{N}_{x_t}(A^{-1}\mathbf{z}_{t+1|t+1}, (A^T(Q + \mathbf{P}_{t+1|t+1})^{-1}A)^{-1})\mathcal{N}_{x_t}(\mu, \Lambda) \tag{D.17}$$

$$\propto \mathcal{N}_{x_t}(m_t, C_t) \tag{D.18}$$

Use Lemma D.0.2 where $\Sigma = Q + \mathbf{P}_{t+1|t+1}$ and $y = \mathbf{z}_{t+1|t+1}$

■ D.0.4 Lemma 4

$$\mathcal{N}_y(Fx + B, R) = C\exp\left[\frac{-1}{2}(y - Fx - B)^T R^{-1}(y - Fx - B)\right] \tag{D.19}$$

$$= C\exp\left[\frac{-1}{2}(F(F^{-1}(y - B) - x))^T R^{-1}F(F^{-1}(y - B) - x)\right] \tag{D.20}$$

$$= C\exp\left[\frac{-1}{2}(F^{-1}(y - B) - x)^T F^T R^{-1}F(F^{-1}(y - B) - x)\right] \tag{D.21}$$

$$= \mathcal{N}_x(F^{-1}(y - B), (F^T R^{-1}F)^{-1}) \tag{D.22}$$

■ D.0.5 Lemma 5

Using Lemma D.0.2 we can solve for

$$N_x(F'(\hat{x})^{-1}(y + F'(\hat{x})\hat{x} - F(\hat{x})), (F'(\hat{x})^T R^{-1} F'(\hat{x}))^{-1})\mathcal{N}_x(\hat{x}, Q) \propto \mathcal{N}_x(m, C) \tag{D.23}$$

Using Lemma D.0.2 where $\mu = \hat{x}$

$$\begin{aligned}
C_t &= Q - QF'(\hat{x})^T(\Sigma + F'(\hat{x})QF'(\hat{x})^T)^{-1}F'(\hat{x})Q \qquad &(\text{D.24})\\
m_t &= \hat{x} + QF'(\hat{x})^T(\Sigma + F'(\hat{x})QF'(\hat{x})^T)^{-1}(y + F'(\hat{x})\hat{x} - F(\hat{x}) - F'(\hat{x})\hat{x}) \qquad &(\text{D.25})\\
&= \hat{x} + QF'(\hat{x})^T(\Sigma + F'(\hat{x})QF'(\hat{x})^T)^{-1}(y - F(\hat{x})) \qquad &(\text{D.26})
\end{aligned}$$

■ D.0.6 Lemma 6

$$\begin{aligned}
&\mathcal{N}_{A^{-1}(x_{t+1}-B)}(z, (A^TQ^{-1}A)^{-1} + P)\\
&= C\exp\left[\frac{-1}{2}(A^{-1}(x_{t+1} - B) - z)^T((A^TQ^{-1}A)^{-1} + P)^{-1}(A^{-1}(x_{t+1} - B) - z)\right] \qquad &(\text{D.27})\\
&= C\exp\left[\frac{-1}{2}(x_{t+1} - B - Az)^T A^{-1T}((A^TQ^{-1}A)^{-1} + P)^{-1}A^{-1}(x_{t+1} - B - Az)\right] \qquad &(\text{D.28})\\
&= \mathcal{N}_{x_{t+1}}(Az + B, (A^{-1T}((A^TQ^{-1}A)^{-1} + P)^{-1}A^{-1})^{-1}) \qquad &(\text{D.29})\\
&= \mathcal{N}_{x_{t+1}}(Az + B, A((A^TQ^{-1}A)^{-1} + P)A^T) \qquad &(\text{D.30})\\
&= \mathcal{N}_{x_{t+1}}(Az + B, A(A^TQ^{-1}A)^{-1}A^T + APA^T) \qquad &(\text{D.31})\\
&= \mathcal{N}_{x_{t+1}}(Az + B, AA^{-1}QA^{-1T}A^T + APA^T) \qquad &(\text{D.32})\\
&= \mathcal{N}_{x_{t+1}}(Az + B, Q + APA^T) \qquad &(\text{D.33})
\end{aligned}$$

Bibliography

[1] Jpl image gallery. `http://www.jpl.nasa.gov/spaceimages/`. Accessed: 2015-04-01.

[2] Nrao image gallery. `https://public.nrao.edu/gallery`. Accessed: 2015-04-01.

[3] B. P. et al. Abbott. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102.

[4] F. Adib and D. Katabi. See through walls with wifi! *ACM*, 43(4):75–86, 2013.

[5] Kazunori Akiyama, Shiro Ikeda, Mollie Pleau, Vincent L Fish, Fumie Tazaki, Kazuki Kuramochi, Avery E Broderick, Jason Dexter, Monika Mośćibrodzka, Michael Gowanlock, et al. Superresolution full-polarimetric imaging for radio interferometry with sparse modeling. *The Astronomical Journal*, 153(4):159, 2017.

[6] Kazunori Akiyama, Kazuki Kuramochi, Shiro Ikeda, Vincent L. Fish, Fumie Tazaki, Mareki Honma, Sheperd S. Doeleman, Avery E. Broderick, Jason Dexter, Monika Mo cibrodzka, Katherine L. Bouman, Andrew A. Chael, and Masamichi Zaizen. Imaging the schwarzschild-radius-scale structure of m87 with the event horizon telescope using sparse modeling. *The Astrophysical Journal*, 838(1):1, 2017.

[7] Steven A Balbus and John F Hawley. A powerful local shear instability in weakly magnetized disks. i-linear analysis. ii-nonlinear evolution. *The Astrophysical Journal*, 376:214–233, 1991.

[8] James M Bardeen. Timelike and null geodesics in the kerr metric. *Black holes*, 215, 1973.

[9] Fabien Baron, John D Monnier, and Brian Kloppenborg. A novel image reconstruction software for optical/infrared interferometry. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 77342I–77342I. International Society for Optics and Photonics, 2010.

[10] Fabien Baron, William D Cotton, Peter R Lawson, Steve T Ridgway, Alicia Aarnio, John D Monnier, Karl-Heinz Hofmann, Dieter Schertl, Gerd Weigelt, Eric Thiébaut, et al. The 2012 interferometric imaging beauty contest. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 84451E–84451E. International Society for Optics and Photonics, 2012.

[11] Harrison H Barrett and Kyle J Myers. *Foundations of image science.* John Wiley & Sons, 2013.

[12] George Keith Batchelor. *An introduction to fluid dynamics.* Cambridge university press, 2000.

[13] J.-P. Berger, F. Malbet, F. Baron, A. Chiavassa, G. Duvert, M. Elitzur, B. Freytag, F. Gueth, S. Hnig, J. Hron, H. Jang-Condell, J.-B. Le Bouquin, J.-L. Monin, J.D. Monnier, G. Perrin, B. Plez, T. Ratzka, S. Renard, S. Stefl, E. Thibaut, K.R.W. Tristram, T. Verhoelst, S. Wolf, and J. Young. Imaging the heart of astrophysical objects with optical long-baseline interferometry. *The Astronomy and Astrophysics Review*, 20(1):53, 2012. ISSN 0935-4956. doi: 10.1007/s00159-012-0053-0.

[14] Christopher M Bishop. *Pattern recognition and machine learning.* springer, 2006.

[15] Abhijit Biswas and Krishnan Mani. Relativistic perihelion precession of orbits of venus and the earth. *Open Physics*, 6(3):754–758, 2008.

[16] P.V. Borges, A. Tews, and D. Haddon. Pedestrian detection in industrial environments: Seeing around corners. 2012.

[17] Katherine L Bouman, Michael D Johnson, Daniel Zoran, Vincent L Fish, Shepherd S Doeleman, and William T Freeman. Computational imaging for vlbi image reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 913–922, 2016.

[18] Geoffrey C Bower, Heino Falcke, Robeson M Herrnstein, Jun-Hui Zhao, WM Goss, and Donald C Backer. Detection of the intrinsic size of sagittarius a* through closure amplitude imaging. *Science*, 304(5671):704–708, 2004.

[19] Geoffrey C Bower, WM Goss, Heino Falcke, Donald C Backer, and Yoram Lithwick. The intrinsic size of sagittarius a* from 0.35 to 6 cm. *The Astrophysical Journal Letters*, 648(2):L127, 2006.

[20] Ronald Bracewell. *Fourier analysis and imaging.* Springer Science & Business Media, 2004.

[21] A. E. Broderick and A. Loeb. Imaging optically-thin hotspots near the black hole horizon of Sgr A* at radio and near-infrared wavelengths. , 367:905–916, April 2006. doi: 10.1111/j.1365-2966.2006.10152.x.

[22] A. E. Broderick, A. Loeb, and R. Narayan. The event horizon of sagittarius a*. *The Astrophysical Journal*, 2009.

[23] A. E. Broderick, V. L. Fish, S. S. Doeleman, and A. Loeb. Evidence for Low Black Hole Spin and Physically Motivated Accretion Models from Millimeter-VLBI Observations of Sagittarius A*. *The Astrophysical Journal*, July 2011.

[24] DF Buscher. Direct maximum-entropy image reconstruction from the bispectrum. In *Very High Angular Resolution Imaging*, pages 91–93. Springer, 1994.

[25] Sean M Carroll. *Spacetime and geometry. An introduction to general relativity*, volume 1. 2004.

[26] Andrew A. Chael, Michael D. Johnson, Ramesh Narayan, Sheperd S. Doeleman, John F. C. Wardle, and Katherine L. Bouman. High-resolution linear polarimetric imaging for the event horizon telescope. *The Astrophysical Journal*, 829(1):11, 2016.

[27] C.-K. Chan, D. Psaltis, F. Özel, R. Narayan, and A. Sadowski. The Power of Imaging: Constraining the Plasma Properties of GRMHD Simulations using EHT Observations of Sgr A*. , 799:1, January 2015. doi: 10.1088/0004-637X/799/1/1.

[28] A. L. Cohen. Anti-pinhole imaging. *Optica Acta: International Journal of Optics*, 29(1):63–67, 1982.

[29] TJ Cornwell and PN Wilkinson. A new method for making maps with unstable radio interferometers. *Monthly Notices of the Royal Astronomical Society*, 196(4): 1067–1086, 1981.

[30] A. Dabbech, C. Ferrari, D. Mary, E. Slezak, O. Smirnov, and J. S. Kenyon. MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators. A sparse deconvolution algorithm for radio interferometric imaging. , 576:A7, April 2015. doi: 10.1051/0004-6361/201424602.

[31] Adrian V Dalca, Katherine L Bouman, William T Freeman, Natalia S Rost, Mert R Sabuncu, and Polina Golland. Population based image imputation. In *International Conference on Information Processing in Medical Imaging*, pages 659–671. Springer, 2017.

[32] Wilbur B Davenport and William L Root. *Random signals and noise.* McGraw-Hill New York, 1958.

[33] A. Davis*, K. L. Bouman*, J. G. Chen, M. Rubinstein, O. Bykztrk, F. Durand, and W. T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):732–745, April 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2016.2622271.

[34] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. Visual vibrometry: Estimating material properties from small motion in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5335–5343, 2015.

[35] S. Doeleman, E. Agol, D. Backer, F. Baganoff, G. C. Bower, A. Broderick, A. Fabian, V. Fish, C. Gammie, P. Ho, M. Honman, T. Krichbaum, A. Loeb, D. Marrone, M. Reid, A. Rogers, I. Shapiro, P. Strittmatter, R. Tilanus, J. Weintroub, A. Whitney, M. Wright, and L. Ziurys. Imaging an Event Horizon: submm-VLBI of a Super Massive Black Hole. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *Astronomy*, 2009.

[36] S. S. Doeleman, J. Weintroub, A. E. E. Rogers, R. Plambeck, R. Freund, R. P. J. Tilanus, P. Friberg, L. M. Ziurys, J. M. Moran, B. Corey, K. H. Young, D. L. Smythe, M. Titus, D. P. Marrone, R. J. Cappallo, D. C.-J. Bock, G. C. Bower,

R. Chamberlin, G. R. Davis, T. P. Krichbaum, J. Lamb, H. Maness, A. E. Niell, A. Roy, P. Strittmatter, D. Werthimer, A. R. Whitney, and D. Woody. Event-horizon-scale structure in the supermassive black hole candidate at the Galactic Centre. , 455:78–80, September 2008. doi: 10.1038/nature07245.

[37] Sheperd S Doeleman, Vincent L Fish, David E Schenck, Christopher Beaudoin, Ray Blundell, Geoffrey C Bower, Avery E Broderick, Richard Chamberlin, Robert Freund, Per Friberg, et al. Jet-launching structure resolved near the supermassive black hole in m87. *Science*, 2012.

[38] Frank W Dyson, Arthur S Eddington, and Charles Davidson. A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 220(571-581): 291–333, 1920.

[39] A. Einstein. Die grundlage der allgemeinen relativittstheorie. *Annalen der Physik*, 354(7):769–822, 1916. ISSN 1521-3889. doi: 10.1002/andp.19163540702.

[40] Marcello Felli and Ralph E Spencer. *Very Long Baseline Interferometry.* Springer, 1989.

[41] R. Fergus, A. Torralba, and W.T. Freeman. Random lens imaging. 2006.

[42] V. Fish, K. Akiyama, K. Bouman, A. Chael, M. Johnson, S. Doeleman, L. Blackburn, J. Wardle, and W. Freeman. Observingand ImagingActive Galactic Nuclei with the Event Horizon Telescope. *Galaxies*, 4:54, October 2016. doi: 10.3390/galaxies4040054.

[43] Vincent L. Fish, Sheperd S. Doeleman, Christopher Beaudoin, Ray Blundell, David E. Bolin, Geoffrey C. Bower, Richard Chamberlin, Robert Freund, Per Friberg, Mark A. Gurwell, Mareki Honma, Makoto Inoue, Thomas P. Krichbaum, James Lamb, Daniel P. Marrone, James M. Moran, Tomoaki Oyama, Richard Plambeck, Rurik Primiani, Alan E. E. Rogers, Daniel L. Smythe, Jason SooHoo, Peter Strittmatter, Remo P. J. Tilanus, Michael Titus, Jonathan Weintroub, Melvyn Wright, David Woody, Ken H. Young, and Lucy M. Ziurys. 1.3mm wavelength vlbi of sagittarius a*: Detection of time-variable emission on event horizon scales. *The Astrophysical Journal Letters*, 727(2):L36, 2011.

[44] Vincent L Fish, Michael D Johnson, Ru-Sen Lu, Sheperd S Doeleman, Katherine L Bouman, Daniel Zoran, William T Freeman, Dimitrios Psaltis, Ramesh Narayan, Victor Pankratius, et al. Imaging an event horizon: Mitigation of scattering toward sagittarius a*. *The Astrophysical Journal*, 2014.

[45] Vincent L. Fish, Michael D. Johnson, Sheperd S. Doeleman, Avery E. Broderick, Dimitrios Psaltis, Ru-Sen Lu, Kazunori Akiyama, Walter Alef, Juan Carlos Algaba, Keiichi Asada, Christopher Beaudoin, Alessandra Bertarini, Lindy Blackburn, Ray Blundell, Geoffrey C. Bower, Christiaan Brinkerink, Roger Cappallo, Andrew A. Chael, Richard Chamberlin, Chi-Kwan Chan, Geoffrey B. Crew, Jason Dexter, Matt Dexter, Sergio A. Dzib, Heino Falcke, Robert Freund, Per Friberg, Christopher H. Greer, Mark A. Gurwell, Paul T. P. Ho, Mareki Honma, Makoto Inoue, Tim Johannsen, Junhan Kim, Thomas P. Krichbaum, James Lamb, Jonathan Len-Tavares, Abraham Loeb, Laurent Loinard, David MacMahon, Daniel P. Marrone, James M. Moran, Monika Mo cibrodzka, Gisela N. Ortiz-Len, Tomoaki Oyama, Feryal zel, Richard L. Plambeck, Nicolas Pradel, Rurik A. Primiani, Alan E. E. Rogers, Katherine Rosenfeld, Helge Rottmann, Alan L. Roy, Chester Ruszczyk, Daniel L. Smythe, Jason SooHoo, Justin Spilker, Jordan Stone, Peter Strittmatter, Remo P. J. Tilanus, Michael Titus, Laura Vertatschitsch, Jan Wagner, John F. C. Wardle, Jonathan Weintroub, David Woody, Melvyn Wright, Paul Yamaguchi, Andr Young, Ken H. Young, J. Anton Zensus, and Lucy M. Ziurys. Persistent asymmetric structure of sagittarius a* on event horizon scales. *The Astrophysical Journal*, 820(2):90, 2016.

[46] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 2015.

[47] Karl Gebhardt, Joshua Adams, Douglas Richstone, Tod R Lauer, SM Faber, Kayhan Gültekin, Jeremy Murphy, and Scott Tremaine. The black hole mass in m87 from gemini/nifs adaptive optics observations. *The Astrophysical Journal*, 729(2):119, 2011.

[48] Reinhard Genzel, Frank Eisenhauer, and Stefan Gillessen. The galactic center massive black hole and nuclear star cluster. *Reviews of Modern Physics*, 82(4): 3121, 2010.

[49] Neil Gershenfeld. *The physics of information technology.* Cambridge University Press, 2000.

[50] A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. R. Morris, S. Yelda, E. E. Becklin, T. Kremenek, M. Milosavljevic, and J. Naiman. Measuring Distance and Properties of the Milky Way's Central Supermassive Black Hole with Stellar Orbits. , 689:1044–1062, December 2008. doi: 10.1086/592738.

[51] Daniel T Gillespie. The mathematics of brownian motion and johnson noise. *American Journal of Physics*, 64(3):225–240, 1996.

[52] VL Ginzburg and SI Syrovatskii. Cosmic magnetobremsstrahlung (synchrotron radiation). *Annual Review of Astronomy and Astrophysics*, 3(1):297–350, 1965.

[53] R. Gold, J. C. McKinney, M. D. Johnson, and S. S. Doeleman. Probing the magnetic field structure in Sgr A* on Black Hole Horizon Scales with Polarized Radiative Transfer Simulations. *ArXiv e-prints*, January 2016.

[54] Richard F Haines and Sherry L Chuang. The effects of video compression on acceptability of images for monitoring life sciences experiments. 1992.

[55] John M Hammersley and Peter Clifford. Markov fields on finite graphs and lattices. 1971.

[56] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, June 2014. doi: 10.1109/CVPR.2014.418.

[57] Gabor T Herman. *Fundamentals of computerized tomography: image reconstruction from projections.* Springer Science & Business Media, 2009.

[58] Gabor T Herman and Joachim Frank. *Computational methods for three-dimensional microscopy reconstruction.* Springer Science & Business Media, 2014.

[59] GT Herman. Image reconstruction from projections: the fundamentals of computerized tomography. 1980. *New York, Academic.*

[60] J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. , 15:417, June 1974.

[61] JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl*, 1974.

[62] M. Honma, K. Akiyama, F. Tazaki, K. Kuramochi, S. Ikeda, K. Hada, and M. Uemura. Imaging black holes with sparse modeling. *Journal of Physics Conference Series*, 699(1):012006, March 2016. doi: 10.1088/1742-6596/699/1/012006.

[63] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[64] Scott A Hughes. Trust but verify: The case for astrophysical black holes. *arXiv preprint hep-ph/0511217*, 2005.

[65] T. Isernia, V. Pascazio, R. Pierri, and G. Schirinzi. Synthetic aperture radar imaging from phase-corrupted data. *IEE Proceedings - Radar, Sonar and Navigation*, Aug 1996. ISSN 1350-2395. doi: 10.1049/ip-rsn:19960458.

[66] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

[67] S Jaeger. The common astronomy software application (casa). In *Astronomical Data Analysis Software and Systems XVII*, 2008.

[68] T. Johannsen and D. Psaltis. Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. II. Black Hole Images. , 718:446–454, July 2010. doi: 10.1088/0004-637X/718/1/446.

[69] M. D. Johnson. Stochastic Optics: A Scattering Mitigation Framework for Radio Interferometric Imaging. , 833:74, December 2016. doi: 10.3847/1538-4357/833/1/74.

[70] M. D. Johnson, V. L. Fish, S. S. Doeleman, D. P. Marrone, R. L. Plambeck, J. F. C. Wardle, K. Akiyama, K. Asada, C. Beaudoin, L. Blackburn, R. Blundell, G. C. Bower, C. Brinkerink, A. E. Broderick, R. Cappallo, A. A. Chael, G. B. Crew, J. Dexter, M. Dexter, R. Freund, P. Friberg, R. Gold, M. A. Gurwell, P. T. P. Ho, M. Honma, M. Inoue, M. Kosowsky, T. P. Krichbaum, J. Lamb, A. Loeb, R.-S. Lu, D. MacMahon, J. C. McKinney, J. M. Moran, R. Narayan, R. A. Primiani, D. Psaltis, A. E. E. Rogers, K. Rosenfeld, J. SooHoo, R. P. J. Tilanus, M. Titus, L. Vertatschitsch, J. Weintroub, M. Wright, K. H. Young, J. A. Zensus, and L. M. Ziurys. Resolved magnetic-field structure and variability near

the event horizon of Sagittarius A*. *Science*, 350:1242–1245, December 2015. doi: 10.1126/science.aac7087.

[71] M. D. Johnson, K. L. Bouman, L. Blackburn, Rosen J. Shiokawa H. Chael A. A, Roelofs F., Akiyama K, Fish V. L., and Doeleman S. S. Dynamical Imaging With Interferometry. *in prep.*, June 2017.

[72] Michael D. Johnson, Abraham Loeb, Hotaka Shiokawa, Andrew A. Chael, and Sheperd S. Doeleman. Measuring the direction and angular velocity of a black hole accretion disk via lagged interferometric covariance. *The Astrophysical Journal*, 813(2):132, 2015.

[73] Svetlana Jorstad and Alan Marscher. The vlba-bu-blazar multi-wavelength monitoring program. *Galaxies*, 4(4):47, 2016.

[74] Svetlana G Jorstad, Alan P Marscher, Matthew L Lister, Alastair M Stirling, Timothy V Cawthorne, Walter K Gear, José L Gómez, Jason A Stevens, Paul S Smith, James R Forster, et al. Polarimetric observations of 15 active galactic nuclei at high frequencies: jet kinematics from bimonthly monitoring with the very long baseline array. *The Astronomical Journal*, 2005.

[75] A. Kadambi, H. Zhao, B. Shi, and R. Raskar. Occluded imaging with time-of-flight sensors, 2016. ACM Transactions on Graphics.

[76] Ayman Bin Kamruddin and Jason Dexter. A geometric crescent model for black hole images. *Monthly Notices of the Royal Astronomical Society*, 434(1):765–771, 2013.

[77] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[78] KI Kellermann, ML Lister, DC Homan, RC Vermeulen, MH Cohen, E Ros, M Kadler, JA Zensus, and YY Kovalev. Sub-milliarcsecond imaging of quasars and active galactic nuclei. iii. kinematics of parsec-scale radio jets. *The Astrophysical Journal*, 609(2):539, 2004.

[79] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific reports*, 6, 2016.

[80] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning.* The MIT Press, 2009. ISBN 0262013193, 9780262013192.

[81] TP Krichbaum, DA Graham, M Bremer, W Alef, A Witzel, JA Zensus, and A Eckart. Sub-milliarcsecond imaging of sgr a* and m 87. In *Journal of Physics: Conference Series.* IOP Publishing, 2006.

[82] M. Laurenzis, A. Velten, and J. Klein. Dual-mode optical sensing: three-dimensional imaging and seeing around a corner. *Optical Engineering*, 2017.

[83] Peter R Lawson, William D Cotton, Christian A Hummel, John D Monnier, Ming Zhao, John S Young, Hrobjartur Thorsteinsson, Serge C Meimon, Laurent M Mugnier, Guy Le Besnerais, et al. An interferometry imaging beauty contest. In *Astronomical Telescopes and Instrumentation.* International Society for Optics and Photonics, 2004.

[84] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[85] R.-S. Lu, A. E. Broderick, F. Baron, J. D. Monnier, V. L. Fish, S. S. Doeleman, and V. Pankratius. Imaging the Supermassive Black Hole Shadow and Jet Base of M87 with the Event Horizon Telescope. *The Astrophysical Journal*, June 2014. doi: 10.1088/0004-637X/788/2/120.

[86] Ru-Sen Lu, Avery E Broderick, Fabien Baron, John D Monnier, Vincent L Fish, Sheperd S Doeleman, and Victor Pankratius. Imaging the supermassive black hole shadow and jet base of m87 with the event horizon telescope. *The Astrophysical Journal*, 788(2):120, 2014.

[87] Ru-Sen Lu, Freek Roelofs, Vincent L. Fish, Hotaka Shiokawa, Sheperd S. Doeleman, Charles F. Gammie, Heino Falcke, Thomas P. Krichbaum, and J. Anton Zensus. Imaging an event horizon: Mitigation of source variability of sagittarius a*. *The Astrophysical Journal*, 817(2):173, 2016.

[88] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. Digital camera identification

from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006.

[89] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine*, 58 (6):1182–1195, 2007.

[90] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001.

[91] Lia Medeiros, Chi-kwan Chan, Feryal Özel, Dimitrios Psaltis, Junhan Kim, Daniel Marrone, and Aleksander Sadowski. Variability in grmhd simulations of sgr a ˆ*: Implications for eht closure phase observations. *arXiv preprint arXiv:1610.03505*, 2016.

[92] John D Monnier and Ronald J Allen. Radio and optical interferometry: Basic observing techniques and data analysis. In *Planets, Stars and Stellar Systems*. Springer, 2013.

[93] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, March 2013. ISSN 2168-6831. doi: 10.1109/MGRS.2013.2248301.

[94] Mocibrodzka, Monika, Falcke, Heino, and Shiokawa, Hotaka. General relativistic magnetohydrodynamical simulations of the jet in m87. *AA*, 586:A38, 2016. doi: 10.1051/0004-6361/201526630.

[95] D. C. Munson, J. D. O'Brien, and W. K. Jenkins. A tomographic formulation of spotlight-mode synthetic aperture radar. *Proceedings of the IEEE*, Aug 1983. ISSN 0018-9219. doi: 10.1109/PROC.1983.12698.

[96] David C Munson, James D O'Brien, and W Kenneth Jenkins. A tomographic formulation of spotlight-mode synthetic aperture radar. *Proceedings of the IEEE*, 71(8):917–925, 1983.

[97] R. Narayan and R. Nityananda. Maximum entropy image restoration in astronomy. , 24:127–170, 1986. doi: 10.1146/annurev.aa.24.090186.001015.

[98] Ezra T Newman and AI Janis. Note on the kerr spinning-particle metric. *Journal of Mathematical Physics*, 6(6):915–917, 1965.

[99] K. Nishino and S.K. Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.

[100] Alan V Oppenheim and Alan S Willsky. *Signals and systems.* Prentice-Hall, 1997.

[101] R. Pandharkar, A. Velten, A. Bardagjy, Bawendi M. Lawson, E., and R. Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments, 2011. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 265-272).

[102] Thomas A Pauls, John S Young, WD Cotton, and JD Monnier. A data exchange standard for optical (visible/ir) interferometry. *Publications of the Astronomical Society of the Pacific*, 117(837):1255–1262, 2005.

[103] TJ Pearson and ACS Readhead. Image formation by self-calibration in radio astronomy. *Annual review of astronomy and astrophysics*, 22(1):97–130, 1984.

[104] Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.

[105] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.

[106] William K Pratt. Generalized wiener filtering computation techniques. *IEEE Transactions on Computers*, 100(7):636–641, 1972.

[107] Hung-Yi Pu, Kazunori Akiyama, and Keiichi Asada. The effects of accretion flow dynamics on the black hole shadow of sagittarius a. *The Astrophysical Journal*, 831(1):4, 2016.

[108] Martin J Rees and Marta Volonteri. Massive black holes: formation and evolution. *Proceedings of the International Astronomical Union*, 2(S238):51–58, 2006.

[109] Stéphanie Renard, Eric Thiébaut, and Fabien Malbet. Image reconstruction in optical interferometry: benchmarking the regularization. *Astronomy & Astrophysics*, 2011.

[110] Freek Roelofs. Studying rapid variability of Sgr A* and M87 with closure phases. Master's thesis, Radboud University, the Netherlands, 2016.

[111] Kristen Rohlfs and Thomas Wilson. *Tools of radio astronomy.* Springer Science & Business Media, 2013.

[112] M Ryle. The new cambridge radio telescope. *Nature*, 194:517–518, 1962.

[113] EJ Schreier, JO Burns, and ED Feigelson. Detection of radio emission from the jet in centaurus a. *The Astrophysical Journal*, 251:523–529, 1981.

[114] Bernard Schutz. *A first course in general relativity.* Cambridge university press, 2009.

[115] MC Shepherd. Difmap: An interactive program for synthesis imaging. In *Astronomical Data Analysis Software and Systems VI*, volume 125, page 77, 1997.

[116] D. Shin, A. Kirmani, V.K. Goyal, and J.H. Shapiro. Computational 3d and reflectivity imaging with high photon efficiency. *Image Processing (ICIP), 2014 IEEE International Conference*, 2014.

[117] D. Shin, A. Kirmani, V.K. Goyal, and J.H. Shapiro. Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors. *IEEE Transactions on Computational Imaging*, 2015.

[118] H. Shiokawa. *General-relativistic magnetohydrodynamics simulations of black hole accretion disks: Dynamics and radiative properties.* PhD thesis, University of Illinois at Urbana-Champaign, 2013.

[119] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Transactions on Graphics (TOG)*, 2016.

[120] John Skilling. Quantified maximum entropy. In *Maximum Entropy and Bayesian Methods.* Springer, 1990.

[121] Greg B Taylor, Chris Luke Carilli, and Richard A Perley. Synthesis imaging in radio astronomy ii. In *Synthesis Imaging in Radio Astronomy II*, 1999.

[122] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[123] Jean-Baptiste Thibault, Ken D Sauer, Charles A Bouman, and Jiang Hsieh. A three-dimensional statistical approach to improved image quality for multislice helical ct. *Medical physics*, 2007.

[124] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr. *Interferometry and Synthesis in Radio Astronomy, 3rd Edition.* 2017. doi: 10.1007/978-3-319-44431-4.

[125] A. Torralba and W. T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. *Computer Vision and Pattern Recognition (CVPR). IEEE.*, pages 374–381, 2012.

[126] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1521–1528. IEEE, 2011.

[127] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391–412, 2003.

[128] C. M. Urry and P. Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. , 107:803, September 1995. doi: 10.1086/133630.

[129] A Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M.G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3(3):745, 2012. ACM Transactions on Graphics.

[130] Jonelle L Walsh, Aaron J Barth, Luis C Ho, and Marc Sarzi. The m87 black hole mass from gas-dynamical models of space telescope imaging spectrograph observations. *The Astrophysical Journal*, 770(2):86, 2013.

[131] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on.* Ieee, 2003.

[132] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA, 1949.

[133] H.Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W.T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *IEEE Signal Processing Letters*, 2012.

[134] L. Xia, C.C. Chen, and J.K. Aggarwal. Human detection using depth information by kinect. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011.

[135] F Xu, D Shin, D Venkatraman, R Lussana, F Villa, F Zappa, V.K. Goyal, F. Wong, and J.H. Shapiro. Photon-efficient computational imaging with a single-photon camera. *Computational Optical Sensing and Imaging*, 2016.

[136] Feng Yuan, Eliot Quataert, and Ramesh Narayan. Nonthermal electrons in radiatively inefficient accretion flow models of sagittarius a*. *The Astrophysical Journal*, 598(1):301, 2003.

[137] Z. Zhang, P. Isola, and E.H. Adelson. Sparklevision: Seeing the world through random specular microfacets. 2014.

[138] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011.

[139] Daniel Zoran and Yair Weiss. Natural images, gaussian mixtures and dead leaves. In *Advances in Neural Information Processing Systems*, 2012.