MIT Open Access Articles

Reconstructing Video of Time-Varying Sources from Radio Interferometric Measurements

Massachusetts Institute of Technology

Reconstructing Video from Interferometric Measurements of Time-Varying Sources

Katherine L. Bouman[1,2], Michael D. Johnson[2], Adrian V. Dalca[1,3], Andrew A. Chael[2],
Freek Roelofs[4], Sheperd S. Doeleman[2], William T. Freeman[1,5]

[1]Massachusetts Institute of Technology, CSAIL [2]Harvard-Smithsonian Center for Astrophysics
[3]Massachusetts General Hospital, HMS [4]Radboud University [5]Google Research

Abstract—**Very long baseline interferometry (VLBI) makes it possible to recover images of astronomical sources with extremely high angular resolution. Most recently, the Event Horizon Telescope (EHT) has extended VLBI to short millimeter wavelengths with a goal of achieving angular resolution sufficient for imaging the event horizons of nearby supermassive black holes. VLBI provides measurements related to the underlying source image through a sparse set spatial frequencies. An image can then be recovered from these measurements by making assumptions about the underlying image. One of the most important assumptions made by conventional imaging methods is that over the course of a night's observation the image is static. However, for quickly evolving sources, such as the galactic center's supermassive black hole (SgrA*) targeted by the EHT, this assumption is violated and these conventional imaging approaches fail. In this work we propose a new way to model VLBI measurements that allows us to recover both the appearance and dynamics of an evolving source by reconstructing a video rather than a static image. By modeling VLBI measurements using a Gaussian Markov Model, we are able to propagate information across observations in time to reconstruct a video, while simultaneously learning about the dynamics of the source's emission region. We demonstrate our proposed Expectation-Maximization (EM) algorithm, StarWarps, on realistic synthetic observations of black holes, and show how it substantially improves results compared to conventional imaging algorithms. Additionally, we demonstrate StarWarps on real VLBI data of the M87 Jet from the VLBA.**

I. INTRODUCTION

Imaging distant astronomical sources with high resolving power at radio wavelengths would require single-dish telescopes with prohibitively large diameters due to the inverse relationship between angular resolution and telescope diameter. Very long baseline interferometry (VLBI) is a technique that 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 [1]. Thus far, VLBI has been primarily used to image sources that are static on the time scale of a day's observation. In this work, we extend the technique's applicability to imaging time-varying sources by reconstructing a video of the source.

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, projected

orthogonal to the direction of the target source [1]. Thus, at a single time, for an array with P telescopes, at most $P(P-1)/2$ 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 allows the array to sample additional spatial frequencies along elliptical paths in the frequency domain [1]. Refer to Figure 1. 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 [2], [3], [4]. 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], [7], [8], [9]. 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 [10]. 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 [11]. 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 [12], [13].

SgrA* is a prime target for the Event Horizon Telescope (EHT) – an international project whose goal is to take the first image of the immediate environment around a black hole [14]. Realizing this goal would not only substantiate the existence of a black holes' event horizon, but also aid in studying general relativity in the strong field regime [15], [16]. Unfortunately, the amount of variation predicted for SgrA* suggests that conventional VLBI imaging techniques will be inappropriate for observations taken by the EHT [4], [11]. 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
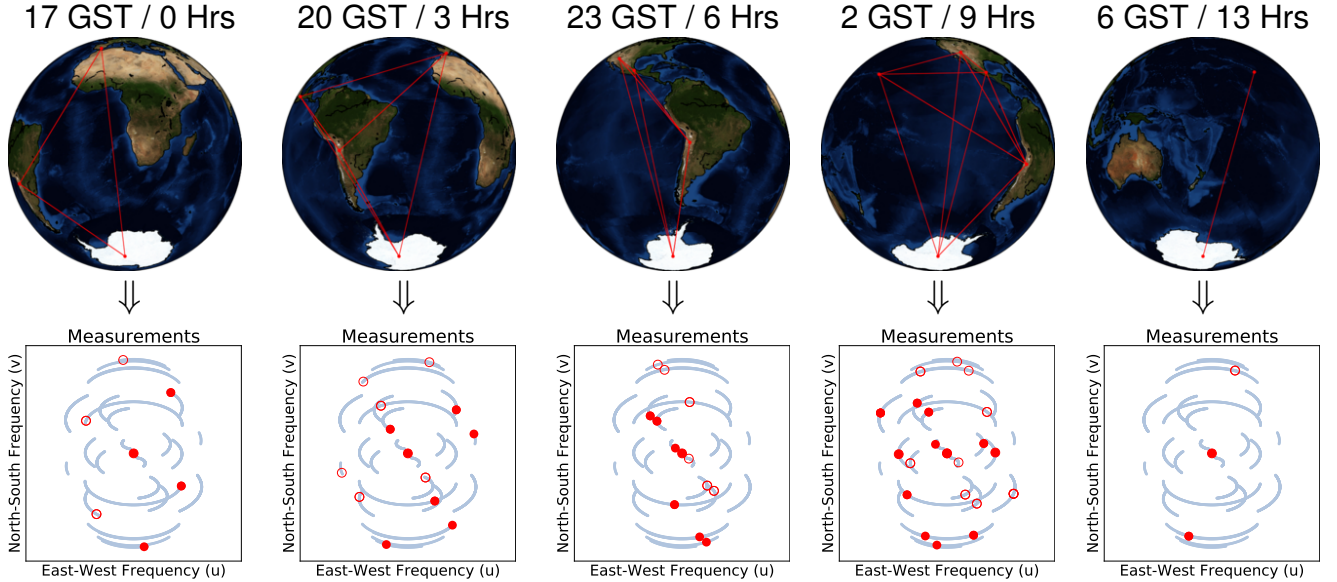
Fig. 1. Earth Rotation Synthesis: For every two telescopes in the interferometric array, we obtain a single measurement (visibility) related to the underlying source image's 2D spatial frequency. This frequency is related to the baseline between the telescopes in the direction perpendicular to the observing source. It is prohibitively difficult to reconstruct a faithful image from just these small number of measurements. 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 progresses (specified by the Greenwich Sidereal Time (GST)), the projected baselines change and the red dots on the frequency coverage plot (bottom row), which indicate 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. The set of all frequencies sampled over an observation is shown by the transparent blue lines. However, when a source evolves over the course of a night's observation, each instantaneous set of measurements (shown in red on each frequency plot) is only related to a single image. 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. The Earth diagrams are shown from the point of view of an observer at Sgr A*. Note that although the telescope array is not changing, the number of visibilities at a given time changes depending on how many sites can still see Sgr A*.

an evolving emission region by simultaneously reconstructing both images and motion trajectories - essentially reconstructing a video rather than a static image.

In Section II we review the basics of interferometric imaging and the data traditionally used to constrain image reconstruction. In Section III we summarize standard approaches to VLBI imaging and define a generalized expression for data consistency. In Section IV we explain how static imaging can be performed under a multivariate Gaussian prior. Highlighting the approach's benefits first in the case of a static source becomes helpful when discussing our proposed solution to the more complex dynamic imaging problem, StarWarps, in Section V, and when deriving our proposed EM inference algorithm in Section VI. Results can be seen in Section VII, followed by concluding remarks in Section VIII.

II. VLBI MEASUREMENTS AND DATA PRODUCTS

A single-dish telescope is diffraction limited, with an angular resolution dictated by the ratio of wavelength to dish diameter [1]. This governing law limits the highest achievable angular resolution of traditional single-dish telescopes. For example, past observations and simulations suggest that the emission surrounding SgrA* subtends approximately 2.5×10^{-10} radians, or 50 μ-arcseconds [15]. Imaging this emission region at 1.3 mm wavelength would require an impossibly large single-dish telescope with a 13,000 km diameter. However, simultaneously collecting data from an array of telescopes, called an interferometer, allows us to overcome the single-dish diffraction limit, and create a virtual telescope as large as

the maximum distance between telescopes in the array. When these telescopes are distributed globally, using separate clocks and recording systems, this technique is referred to as Very Long Baseline Interferometry (VLBI) [1].

An interferometer consists of P telescopes simultaneously observing and recording time-stamped data-streams of light traveling from a common source [1]. From these measurements a number of different *data products* can be computed. 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 paper.

A. Visibilities

The time-averaged cross-correlation of the recorded scalar electric fields at two telescopes, called a *visibility*, provides information about the spatial structure of the observed source[1]. Formally, the van Cittert-Zernike theorem states that a visibility, $\Gamma(I, u, v)$ is related to the ideal source image $I(\varrho, \delta)$ through a Fourier transform:

$$\Gamma(I, u, v) \approx \int_\varrho \int_\delta e^{-i2\pi(u\varrho + v\delta)} I(\varrho, \delta) d\varrho d\delta \qquad (1)$$

where (ϱ, 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 source direction [1]. Note that each visibility is a complex value with both

[1]Although the particular length of time-averaging varies per experiment, for the EHT data is averaged over sub-second intervals before additional coherent averaging is done to increase SNR (on an order of 5-100 seconds).

Visibility Amplitude
ALMA-SPT Baseline

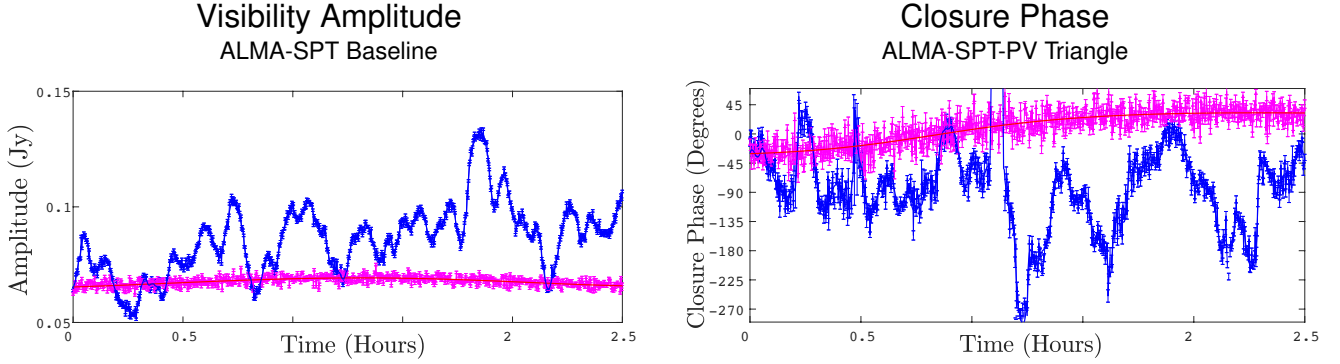## Closure Phase
ALMA-SPT-PV Triangle

Fig. 2. **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 at a sampling interval of 11 seconds using the EHT2017 array from the frames in Video 3 presented in Section VII. This figure shows 1 of the $\binom{K}{2}$ possible visibility amplitudes and 1 of $\binom{K}{3}$ possible closure phases (phase of the bispectrum).

an amplitude and a phase. As each visibility is calculated from a pair of telescopes, for a P telescope array we can obtain up to $\frac{P(P-1)}{2}$ visibility measurements at a single instant in time. The image spatial frequency sampled by a visibility is related to the distance between the telescopes; telescopes that are father apart inform us about higher spatial frequencies of the underlying image. For radio wavelengths, the thermal noise appearing on perfectly calibrated visibilities is circularly Gaussian in the real-imaginary plane[2] [3].

B. Bispectrum and Closure Phases

The van Cittert-Zernike theorem (Equation 1) is derived under the assumption that light is moving through a vacuum. However, in reality, inhomogeneity in the Earth's atmosphere causes the light to move at different velocities towards each telescope. At short observing wavelengths, this propagation delay has a large effect on the phase of measurements, and renders the absolute visibility phase unusable for image reconstruction [17]. Although absolute phase measurements cannot be used, a property called phase closure allows us to still obtain some information from these phases.

Consider three telescopes, denoted by i, j, and k, observing the same source. From each pair of telescopes we compute a visibility: $\Gamma_{i,j}, \Gamma_{j,k}, \Gamma_{k,i}$. Rapidly changing atmospheric propagation delays, resulting in phase errors of ϕ_i and ϕ_j to telescopes i and j respectively, will cause error in the visibility's phase:

$$\Gamma_{i,j}^{\text{meas}} = e^{i(\phi_i - \phi_j)} \Gamma_{i,j}^{\text{ideal}}. \tag{2}$$

However, by multiplying the visibilities from 3 different telescopes in a closed loop, we obtain a term that is invariant to the atmosphere (Equation 3) [18].

$$\Gamma_{i,j}^{\text{meas}} \Gamma_{j,k}^{\text{meas}} \Gamma_{k,i}^{\text{meas}} = e^{i(\phi_i - \phi_j)} \Gamma_{i,j}^{\text{ideal}} e^{i(\phi_j - \phi_k)} \Gamma_{j,k}^{\text{ideal}} e^{i(\phi_k - \phi_i)} \Gamma_{k,i}^{\text{ideal}}$$
$$= \Gamma_{i,j}^{\text{ideal}} \Gamma_{j,k}^{\text{ideal}} \Gamma_{k,i}^{\text{ideal}}. \tag{3}$$

We refer to the multiplication of these three complex visibilities in a closed loop as the *bispectrum*, and the phase of the bispectrum as the *closure phase*. Although these data products are invariant to atmospheric noise, they come at the

---

[2]In practice, the thermal noise is calculated from first principles using measured SEFD values at each telescope (see Eq 42) or estimated empirically.

cost of having a fewer number of independent data points - only $\frac{(P-1)(P-2)}{2}$ compared to $\frac{P(P-1)}{2}$ independent visibilities. Additionally, when using these data products, information related to the absolute location of the source is lost.

As each bispectrum (Eq. 3) is the multiplication of three visibilities with Gaussian noise, its noise is not Gaussian. Although the noise cannot be fully expressed simply with a covariance matrix, we follow [5] and approximate the bispectrum's distribution as Gaussian with a best-fit covariance matrix. Refer to the supplemental material for a derivation of this approximation.

C. Visibility Amplitudes

Bispectrum measurements not only entangle the phase, but also the amplitudes of the three corresponding spatial frequency components. Although atmospheric inhomogeneity causes substantial phase errors in each complex visibility, with careful calibration the amplitude of the visibility can be well estimated. Thus, we have found that during image reconstruction it is quite helpful to additionally constrain the visibility amplitude in addition to the bispectrum or closure phase. Empirically we have found this to be especially important when reconstructing an image with a large field of view. As thermal noise is isotropic in the real-imaginary plane, a perfectly gain-calibrated visibility amplitude has the same standard deviation of noise as a perfectly calibrated complex visibility [1].

III. PREVIOUS IMAGING APPROACHES

VLBI image reconstruction has similarities with other spectral image reconstruction problems, such as Synthetic Aperture Radar (SAR), Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and seismic interferometry [19], [20], [21], [22]. However, 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. Additionally, the frequency coverage obtained in VLBI is dictated by where there are existing telescopes, and often contains large gaps of unsampled space. Accounting for these differences is crucial.

Traditional interferometric imaging methods, such as CLEAN, are not inherently capable of handing the atmosphere's corruption of the visibility phases in mm/sub-mm wavelengths, even for static sources [2], [3]. Thus, they frequently have trouble imaging with sparse, short-wavelength interferometric arrays like the EHT. However, Bayesian-style methods have made it possible to handle sparse and heterogeneous arrays in the mm/sub-mm regime, and can often even achieve accurate super-resolved images in the case of a static source [23], [24], [6].

In order to reconstruct an image from the observed VLBI data, $\boldsymbol{y}$, these methods first approximate the continuous image $I(\varrho, \delta)$ as a square $M \times M$ array of values, $\boldsymbol{x}$, that represents the source image's flux over a specified field of view (FOV)[3]. Using this representation, the imaging methods aim to solve

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left[\chi(\boldsymbol{x}, \boldsymbol{y}) - \gamma \mathcal{R}(\boldsymbol{x}) \right] \qquad (4)$$

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 impact in this minimization process is specified with the hyperparameter γ. Equation 4 can be interpreted probabilistically when $-\chi(\boldsymbol{x}, \boldsymbol{y}) = \log p(\boldsymbol{y}|\boldsymbol{x})$, $\mathcal{R}(\boldsymbol{x}) = \log p(\boldsymbol{x})$, and $\gamma = 1$. In this case, the minimizing cost function is a log-posterior with maximum a posteriori (MAP) estimate $\hat{\boldsymbol{x}}$, and can be solved using a Bayesian inference method. While many methods do not have a probabilistic interpretation, their formulation leads to a similar optimization.

Multiple algorithms have taken this Bayesian-style approach to VLBI imaging. These algorithms often define a similar data inconsistency measure, $\chi(\boldsymbol{x}, \boldsymbol{y})$ (refer to Section III-A), but vary in what characterizes a "good" image, $\mathcal{R}(\boldsymbol{x})$. Maximum entropy, total variation, sparsity, and patch-based priors have all been used to construct $\mathcal{R}(\boldsymbol{x})$, and have been demonstrated in imaging both optical and radio interferometry data taken with sparse, heterogeneous arrays of static sources [5], [6], [8], [25], [26], [27], [28].

VLBI data taken from an evolving source are significantly different from that of a static source [29], [30]. For example, Figure 2 shows a simulated visibility amplitude and closure phase expected for a source evolving over time, and compares it to the data products expected if the same source were static. As the data can no longer be explained by a single image with similar structure, most static imaging methods break down on this data and are unable to produce accurate results.

Recent work has attempted to recover a single average image by first normalizing the data and then smoothing it within a characteristic time-frame [30][4]. 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 VII).

In [31] we simultaneously developed an alternative method for time-variability imaging that also reconstructs a video of the underlying emission region from sparse interferometric measurements. In that work, we develop a more flexible framework with fewer model constraints. However, this modeling choice leads to a much more difficult optimization problem that is prone to local minima, and often requires ad hoc methods to achieve satisfactory convergence. This frequently leads to a solution inconsistent with the true structure of the source images when the data is especially sparse or noisy. Thus, the strengths and weaknesses of this alternative method are complementary to those of the approach we develop here. Similarly, in [32] a Kalman Filter based approach was proposed for video generation from sparse interferometric measurements. This early work has striking similarities to parts of our own presented approach; however, the method was not demonstrated on interferometric data – real or synthetic – and does not address many of the limiting characteristics of EHT-quality data.

The problem of reconstructing video of a moving subject from sparse frequency samples appears in the MRI literature and is often referred to as Dynamic MRI (dMRI) [33], [34], [35], [36]. Recent dMRI approaches reconstruct time-evolving MRI images using a variety of sophisticated regularization techniques, such as non-linear manifold [37] and sparsity models [38], [39]. Although this problem has similarities to VLBI, there are substantial differences that make it not possible to blindly apply approaches developed for dMRI to VLBI. As discussed in Section III-A, due to atmospheric inhomogeneity, VLBI measurements for short wavelength observations lose absolute phase information. Not only does this mean it is not possible to simply use a linear DTFT relationship to recover the underlying image from the data, but additionally the absolute position of the image on the sky is lost. Therefore, the common strawman approach of using a sliding-window to share data across time results in a series of frames with no spatial consistency, leading to a flickering video.

A. Data Consistency

Methods often constrain the image reconstruction using a different set of data products. However, most can be generalized through a common definition of $\chi(\boldsymbol{x}, \boldsymbol{y})$. This generalization will become helpful in defining models in Sections IV and V.

Let $\boldsymbol{y}$ be a K-dimensional real vector of, possibly heterogeneous, data products. We define a function $f(\boldsymbol{x})$ to return the expected value of $\boldsymbol{y}$ if $\boldsymbol{x}$ were the true underlying image. This measurement function, $f(\boldsymbol{x})$, is composed of a set of sub-functions, $g_k(\boldsymbol{x})$, that each simulate an ideally observed data product:

$$f(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ g_2(\boldsymbol{x}) \\ ... \\ g_K(\boldsymbol{x}) \end{bmatrix} : \Re^{M^2 \times 1} \to \Re^{K \times 1}. \qquad (5)$$

Although interferometric observables (e.g. visibilities, bispectrum) are in general complex valued, we break up these

---

[3]The FOV for a time-varying source can be estimated from the variation in visibilities obtained over simultanoulsy observed frequency bands.

[4]In the case of no atmospheric error the visibilities are smoothed. In the case of atmospheric error the bispectrum and visibility amplitudes are smoothed.

complex quantities into their real and imaginary parts in the vector $\boldsymbol{y}$. For example, if we wish to constrain image reconstruction with complex visibilities for P observing telescope sites, then $\boldsymbol{y}_{2k} = \Re\left[\Gamma(u_k, v_k)\right]$ and $\boldsymbol{y}_{2k+1} = \Im\left[\Gamma(u_k, v_k)\right]$ for $K = P(P-1)/2$. In this case,

$$g_{2k}(\boldsymbol{x}) = \Re[\mathbf{f}(u_k, v_k)]^T \boldsymbol{x} \qquad (6)$$

$$g_{2k+1}(\boldsymbol{x}) = \Im[\mathbf{f}(u_k, v_k)]^T \boldsymbol{x}, \qquad (7)$$

for the complex row-vector $\mathbf{f}(u_k, v_k)$ that extracts a single spatial frequency (u_k, v_k) from the vectorized image $\boldsymbol{x}$.

Using these functions, we define $\chi(\boldsymbol{x}, \boldsymbol{y})$ as

$$\chi(\boldsymbol{x}, \boldsymbol{y}) \propto \sum_{k=1}^{K} \frac{(\boldsymbol{y}[k] - g_k(\boldsymbol{x}))^2}{\sigma[k]^2} \qquad (8)$$

$$= (\boldsymbol{y} - f(\boldsymbol{x}))^T \boldsymbol{R}^{-1}(\boldsymbol{y} - f(\boldsymbol{x})), \qquad (9)$$

for $\boldsymbol{R} = \mathrm{diag}[\sigma[1]^2, .., \sigma[K]^2]$ composed of the variance of noise on each $\boldsymbol{y}[k]$. If $\boldsymbol{y}$ is sampled from a Gaussian distribution with mean $f(\boldsymbol{x})$ and covariance $\boldsymbol{R}$ then $\frac{1}{K}\chi(\boldsymbol{x}, \boldsymbol{y}) \approx 1$.

Depending on the data products selected for $\boldsymbol{y}$, the measurement function, $f(\boldsymbol{x})$, may be a linear or non-linear function of $\boldsymbol{x}$. Visibilities can be extracted by performing a linear matrix operation similar to a Discrete Time Fourier Transform (DTFT): $f(\boldsymbol{x}) = \boldsymbol{F}\boldsymbol{x}$ [5]. However, in the presence of atmospheric noise, $\boldsymbol{y}$ may be populated with data products that are invariant to atmospheric inhomogeneity. The bispectrum, closure phase, and visibility amplitude data products are invariant to this error, but come at the cost of requiring a non-linear measurement function, $f(\boldsymbol{x})$ [5], [6], [26].

IV. STATIC MODEL & INFERENCE

Before discussing our proposed approach to dynamic imaging for time-varying sources, we first review interferometric imaging for a static source and discuss a simple, yet instructive, approach using a multivariate Gaussian image prior. The intent of this section is *not* to present a novel and competitive static imaging method, but instead to set up the tools necessary to easily understand dynamic imaging in Sections V and VI.

We measure a vector of real values $\boldsymbol{y}$ that are generated by observing a static source's emission region image, $I(\varrho, \delta)$. These measurements are extremely sparse and noisy, and thus do not fully characterize the underlying image. For example, a simple PCA analysis on rows of the DTFT matrix $\boldsymbol{F}$ for the EHT 2017 campaign (see the uv-coverage of Fig. 4) shows that 95% of the variance can be described using only 1624 measurement sub-functions, $g(\boldsymbol{x})$, for a 10000 pixel image; essentially $\boldsymbol{F}$ constrains only 16% of the unknowns. To solve this problem, we impose a prior distribution on $\boldsymbol{x}$ and seek a maximum a posteriori (MAP) estimate of the underlying image given these sparse observations. We adopt the model presented in [5] to represent $I(\varrho, \delta)$ as vectorized coefficients, $\boldsymbol{x}$. Using this representation, we define our observation model as:

$$\boldsymbol{y} \sim \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}), \qquad (10)$$

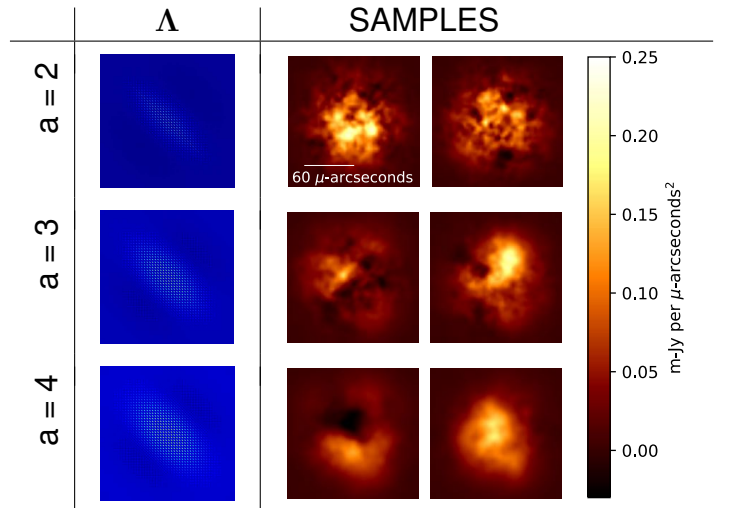$$\boldsymbol{x} \sim \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}), \qquad (11)$$

Fig. 3. **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 (i.e., the prior encourages smoother structure). In these examples μ is a 2D Gaussian image with standard deviation of 75 μ-arcseconds. and $c = 0.5$.

where $\mathcal{N}_z(m, \Sigma)$ is the multivariate normal distribution of z with mean m and covariance Σ. In the chosen model, both the data likelihood, $p(\boldsymbol{y}|\boldsymbol{x})$, and the underlying image prior, $p(\boldsymbol{x})$, are multivariate normal distributions. The posterior probability is written in terms of these two terms:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \qquad (12)$$

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

Unfortunately, $p(\boldsymbol{y}|\boldsymbol{x})$ is not truly Gaussian when $\boldsymbol{y}$ is composed of bispectra or closure phases, as each visibility is used to compute multiple terms. However, as discussed in Section II-B, we assume that each term of $\boldsymbol{y}$ is independent and can be described with a Gaussian noise model. This approximation has been shown to be a good approximation in practice (see the supplemental material) [1], [5].

A. Multivariate Gaussian Image Prior

A prior distribution on $\boldsymbol{x}$ constrains 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], [7], [5], [4]. 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 less sharp image reconstructions 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 VI).

Studies have shown that the average power spectrum of an image often falls with the inverse of spatial frequency in the form $1/(u^2 + v^2)^{a/2}$, where a is a value that specifies the smoothness of the image [40]. 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
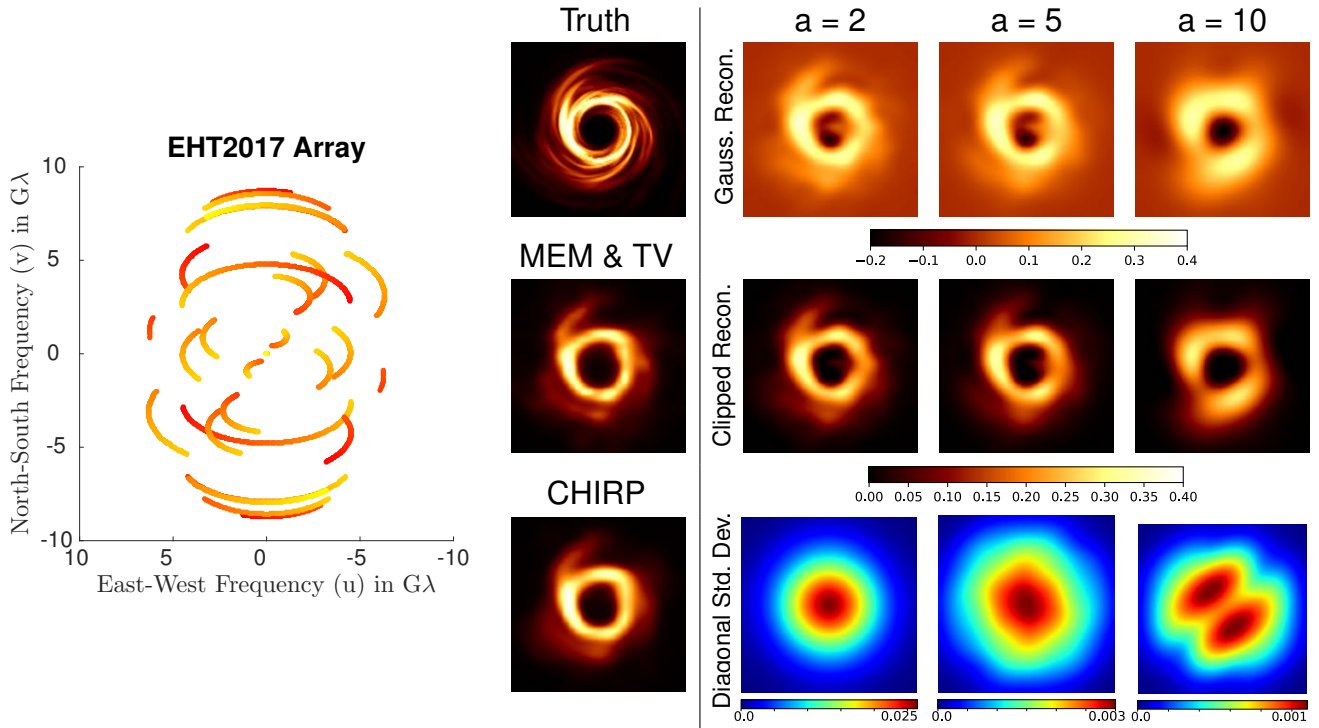
Fig. 4. Static Imaging Comparison: Results of static imaging using a multivariate Gaussian prior (a = 2, 5, 10) compared to state-of-the-art reconstruction methods using MEM & TV regularizers [6] as well as patch-based regularizers (CHIRP) [5]. All images are shown with a field of view of 160 μ-arcseconds. Data is generated using a static image with the uv-coverage of the EHT2017 array shown on the left (see Section VII). The uv-coverage is colored by time, as indicated by the colorbar in Figure 7. Note however that in this static imaging case the time of measurements is not relevant. 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 scaling of the colormaps is in mili-Jansky per squared μ-arcsecond.

a prior distribution. Specifically,

$$\boldsymbol{\Lambda}' = \boldsymbol{W}^{*T}\text{diag}\,[\boldsymbol{b}]\,\boldsymbol{W} \tag{13}$$

$$b[i] = \begin{cases} (u[i]^2 + v[i]^2)^{-a/2} & u[i]^2 + v[i]^2 > 0 \\ \epsilon & u[i]^2 + v[i]^2 = 0 \end{cases} \tag{14}$$

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

$$S = \left\{\frac{m - M/2}{FOV}\right\}, m \in \mathbb{Z} : m \in [0, M-1], \tag{15}$$

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 $c \cdot \text{diag}[\boldsymbol{\mu}]$:

$$\boldsymbol{\Lambda} = c^2 \text{diag}[\boldsymbol{\mu}]^T \boldsymbol{\Lambda}' \text{diag}[\boldsymbol{\mu}] \tag{16}$$

A c value of 1/3 implies that 99% of flux values sampled from $\mathcal{N}_x(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ will be positive. In this work we have chosen $\boldsymbol{\mu}$ to be a circular Gaussian with a standard deviation of 40-50% of the reconstructed FOV to encourage most of the flux to stay near the center of the image and away from the edges.

Figure 3 shows the covariance matrix constructed for $a = 2, 3, 4$ 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.

B. Inference

Our goal is to find the most likely image, $\boldsymbol{x}$, that describes the data products we have observed, $\boldsymbol{y}$. A maximum a posteriori (MAP) solution is found by maximizing the log-posterior from Equation 12:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{argmax}} \log p(\boldsymbol{x}|\boldsymbol{y}) \tag{17}$$

$$= \underset{\boldsymbol{x}}{\text{argmin}} \left[(f(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{R}^{-1}(f(\boldsymbol{x}) - \boldsymbol{y}) \right.$$
$$\left. + (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right]. \tag{18}$$

Note the similarities of this equation's structure to that of previous static imaging methods in Equation 4 and 9. Although the hyperparameter γ is no longer explicit, the scaling of $\boldsymbol{\Lambda}$ acts like this hyperparameter and balances influence of the measured data with influence of the prior.

1) Linear Measurements: As explained in Section III-A, $f(\boldsymbol{x})$ is linear when $\boldsymbol{y}$ is composed solely of calibrated

complex visibilities with no atmospheric error. In this case – when $f(\boldsymbol{x}) = \boldsymbol{F}\boldsymbol{x}$ – a closed-form solution of $\hat{\boldsymbol{x}}$ can be found through traditional Weiner filtering. Specifically, we can compute the most likely estimate of each $\boldsymbol{x}$ as:

$$\hat{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}). \quad (19)$$

In the limit of having no prior information about the underlying image $\boldsymbol{x}$, e.g., $\boldsymbol{\Lambda} = \lim_{\lambda\to\infty} \lambda\mathbb{1}$ for identity matrix $\mathbb{1}$, this MAP solution reduces to $\hat{\boldsymbol{x}} = \boldsymbol{F}^+\boldsymbol{y}$. 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 $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$. This is very similar to reconstructing the "dirty image" [3].

The same solution can also be obtained by evaluating the posterior distribution. With a linear measurement function, $f(\boldsymbol{x})$, the proposed Gaussian formulation leads to a closed-form expression for the posterior. In particular,

$$p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}_{\boldsymbol{x}}(\hat{\boldsymbol{x}}, \boldsymbol{C}). \quad (20)$$

for covariance matrix

$$\boldsymbol{C} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{F}^T(\boldsymbol{R} + \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T)^{-1}\boldsymbol{F}\boldsymbol{\Lambda}. \quad (21)$$

Estimating uncertainty with the covariance matrix is useful in understanding what regions of the reconstructed image we trust. This becomes especially helpful when propagating information in dynamical imaging, as will be demonstrated in Figure 9.

2) 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. However, 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 approximation around $\tilde{\boldsymbol{x}}$, we approximate the data likelihood as

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}_{\boldsymbol{y}}(f(\boldsymbol{x}), \boldsymbol{R}) \approx \mathcal{N}_{\boldsymbol{y}}\left(f(\tilde{\boldsymbol{x}}) + \dot{\boldsymbol{F}}(\boldsymbol{x} - \tilde{\boldsymbol{x}}), \boldsymbol{R}\right), \quad (22)$$

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

$$\hat{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\dot{\boldsymbol{F}}^T(\boldsymbol{R} + \dot{\boldsymbol{F}}\boldsymbol{\Lambda}\dot{\boldsymbol{F}}^T)^{-1}(y - f(\tilde{\boldsymbol{x}}) + \dot{\boldsymbol{F}}(\tilde{\boldsymbol{x}} - \boldsymbol{\mu})). \quad (23)$$

To further improve the solution, we solve Equation 23 iteratively by updating $\hat{\boldsymbol{x}}$ 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 23 reduces to Equation 19. We compare results of this reconstruction method to other state-of-the-art methods for a static source in Figure 4. Figure 4 demonstrates that, although this approach does not outperform other state-of-the-art static imaging methods, reasonable results are achieved despite a simpler image regularizer and optimization procedure. This simpler approach will become useful in developing a dynamic imaging approach.

V. Dynamic Model

Earth rotation synthesis inherently assumes that the source being imaged is static over the course of an observation [3]. If
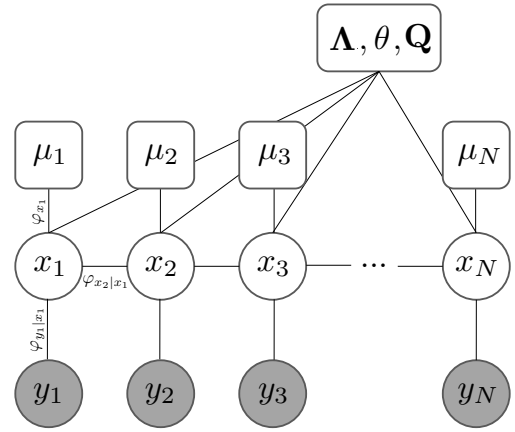

Fig. 5. **Graphical Representation of our Dynamic Imaging Model**: At each time t we observe a vector of data products $\boldsymbol{y}_t$ corresponding to the instantaneous source image $\boldsymbol{x}_t$. We assume each image $\boldsymbol{x}_t$ is related to its adjacent neighbors in time, $\boldsymbol{x}_{t-1}$ 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 constrained by the covariance matrix $\boldsymbol{Q}$. In this diagram, squares indicate parameters, circles are variables, and shaded circles indicate the variable is observed.

this assumption holds, it is possible to collect more than $P(P-1)/2$ measurements that inform us about the underlying source through earth rotation synthesis However, in the case of an evolving source, as is predicted to be the case for SgrA*, this assumption is violated – measurements taken at different times throughout the observation correspond to different underlying source images.

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 5):

$$\varphi_{\boldsymbol{y}_t|\boldsymbol{x}_t} = \mathcal{N}_{\boldsymbol{y}_t}(f_t(\boldsymbol{x}_t), \boldsymbol{R}_t), \quad (24)$$

$$\varphi_{\boldsymbol{x}_t} = \mathcal{N}_{\boldsymbol{x}_1}(\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t), \quad (25)$$

$$\varphi_{\boldsymbol{x}_t|\boldsymbol{x}_{t-1}} = \mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q}), \quad (26)$$

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

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 (26) that describes how images evolve over time. If we assume that there is no evolution between neighboring images in time ($\boldsymbol{A} = \mathbb{1}, \boldsymbol{Q} = \boldsymbol{0}$), this dynamic model reduces to that of static imaging. Using the Hammersley-Clifford Theorem [41], the joint distribution of this dynamic model can be written as a product of its potential functions:

$$p(\mathbf{X}, \mathbf{Y}; \boldsymbol{A}) \propto \prod_{t=1}^{N} \varphi_{\boldsymbol{y}_t|\boldsymbol{x}_t} \prod_{t=1}^{N} \varphi_{\boldsymbol{x}_t} \prod_{t=2}^{N} \varphi_{\boldsymbol{x}_t|\boldsymbol{x}_{t-1}}. \quad (27)$$

A. 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, whereas $\boldsymbol{A} = 2\mathbb{1}$ indicates that the image's brightness doubles at each time step. Since the evolution matrix $\boldsymbol{A}$ is not time dependent, the underlying source image evolves similarly over the entire observation. However, at each time the image can deviate slightly from this persistent evolution. The amount of allowed intensity deviation is expressed in the time-invariant covariance matrix $\boldsymbol{Q}$.

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 $(\boldsymbol{\varrho}, \boldsymbol{\delta})$. These pulses can shift around, causing motion in the image. As described by the Shift Theorem, the shift of a pulse by Δ will change the phase of its Fourier Transform by $2\pi f \Delta$ for each frequency f. Thus, under small motions, we can write $\boldsymbol{A}$ in terms of the image's full $M^2 \times M^2$ DFT matrix $\boldsymbol{W}$ (see Section IV-A), and a column-vector of M pixel shifts: $\boldsymbol{s} = (\boldsymbol{s}_{\boldsymbol{\varrho}}, \boldsymbol{s}_{\boldsymbol{\delta}})$:

$$\boldsymbol{A} = \Re \left[\boldsymbol{W}^{*T} \left(\exp\left[-i2\pi(\boldsymbol{u}\boldsymbol{s}_\varrho^T + \boldsymbol{v}\boldsymbol{s}_\delta^T) \right] \odot \boldsymbol{W} \right) \right]. \quad (28)$$

Applying $\boldsymbol{A}$ to a vectorized image $\boldsymbol{x}_t$ of light pulses results in a new image, $\boldsymbol{x}_{t+1}$, where the pulses have been shifted according to $\boldsymbol{s}$ and re-interpolated on the 2D DFT grid. Note that in the case $\boldsymbol{s} = \boldsymbol{0}$ then $\boldsymbol{A} = \boldsymbol{W}^{*T}\boldsymbol{W} = \mathbb{1}$.

The above parameterization of evolution matrix $\boldsymbol{A}$ in terms of $\boldsymbol{s}$ allows for independent, 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 $\mathcal{M} = \left[\mathcal{M}_\varrho, \mathcal{M}_\delta \right]^T$ of size $2M^2 \times D + 1$, and restrict the motion at every time step to be a linear function of this motion:

$$\begin{bmatrix} \boldsymbol{\varrho}_{t+1} \\ \boldsymbol{\delta}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varrho}_t \\ \boldsymbol{\delta}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{s}_{\varrho_t} \\ \boldsymbol{s}_{\delta_t} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{\varrho_t} \\ \mathcal{M}_{\delta_t} \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix}. \quad (29)$$

This parameterization allows us to describe a wide variety of motion (or warp) fields. Generic "smooth" warp fields can be described by using a truncated Discrete Cosine Transform (DCT) basis as $\mathcal{M}$. However, more compressed motion bases can also be used [42], [43]. In this work, results are shown using an affine transformation parametrized with a four-dimensional θ that captures rotation, shear, and scaling. As an affine transformation, θ, acting on a pulse at location (ϱ_t, δ_t) results in moving the pulse to location

$$\begin{bmatrix} \varrho_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix} \begin{bmatrix} \varrho_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} \varrho_t & \delta_t & 0 & 0 \\ 0 & 0 & \varrho_t & \delta_t \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}, \quad (30)$$

in this work we define,

$$\begin{bmatrix} \mathcal{M}_{\varrho_t} \\ \mathcal{M}_{\delta_t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \varrho_t & \delta_t & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \varrho_t & \delta_t \end{bmatrix}. \quad (31)$$

For example, using this motion basis with $\theta_1 = \cos\phi$, $\theta_2 = \sin\phi$, $\theta_3 = -\sin\phi$, and $\theta_4 = \cos\phi$ would specify that every time step the image is rotated by ϕ radians.

VI. DYNAMIC IMAGING 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 VI-A 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 VI-B, 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.

A. 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, $\boldsymbol{y}$. In the proposed dynamic model, a similar closed-form solution exists in the case of a linear $f(\boldsymbol{x})$ and diagonal $\boldsymbol{R}_t$ and $\boldsymbol{Q}$ matrices [35]. However, this requires us to invert a large $M^4 \times N^2$ non-block-diagonal matrix. Thus, 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 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 [44]. 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 2, 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. By combining these terms we obtain

$$p(\boldsymbol{x}_t|\boldsymbol{Y}) = \mathcal{N}_{x_t}(\hat{\boldsymbol{x}}_t, \boldsymbol{C}_t) \quad (32)$$
$$\propto \mathcal{N}_{x_t}(\mathbf{z}_{t|t-1}^\alpha, \mathbf{P}_{t|t-1}^\alpha)\mathcal{N}_{x_t}(\mathbf{z}_{t|t}^\beta, \mathbf{P}_{t|t}^\beta),$$

Algorithm 1: Forward Updates $t = 1 \rightarrow 2 \rightarrow \dots \rightarrow N$

Predict:

$$\mathbf{z}^\alpha_{t|t-1} = \boldsymbol{A}\mathbf{z}^\alpha_{t-1|t-1}$$
$$\mathbf{P}^\alpha_{t|t-1} = \boldsymbol{Q} + \boldsymbol{A}\mathbf{P}^\alpha_{t-1|t-1}\boldsymbol{A}^T$$

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

Update:

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

Initialization:

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

---

which, as shown in the supplemental material, has mean $\hat{\boldsymbol{x}}_t$ and covariance $\mathbf{C_t}$:

$$\hat{\boldsymbol{x}}_t = \mathbf{P}^\beta_{t|t}(\mathbf{P}^\alpha_{t|t-1} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\alpha_{t|t-1} + \mathbf{P}^\alpha_{t|t-1}(\mathbf{P}^\alpha_{t|t-1} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{z}^\beta_{t|t}$$
$$\mathbf{C_t} = \mathbf{P}^\beta_{t|t-1}(\mathbf{P}^\alpha_{t|t-1} + \mathbf{P}^\beta_{t|t})^{-1}\mathbf{P}^\beta_{t|t}, \tag{33}$$

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 IV-B, 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 [44]. In fact, removing the $\varphi_{\boldsymbol{x}_t}$ term for $t > 1$ in Equation 27 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 $f_t(\boldsymbol{x})$) [32]. 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.

B. Unknown Evolution

If the evolution matrix $\boldsymbol{A}$ is unknown, it is not possible to simply solve for $\boldsymbol{X}$ in the way outlined in Section VI-A.

Instead we choose to use an Expectation-Maximization (EM) algorithm to recover $\boldsymbol{A}$ (parameterized by θ), and then subsequently use the procedure presented in Section VI-A to recover $\boldsymbol{X}$.

The EM algorithm defines an iterative process that solves for the evolution parameters θ that maximize the complete likelihood in Equation 27 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:

- Expectation step (E step): Calculate the expected value of the log likelihood function (see Equation 27), with respect to the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ under the current estimate of the θ parameters, $\theta^{(i)}$:

$$Q(\theta|\theta^{(i)}) = E_{\mathbf{X}|\mathbf{Y},\theta^{(i)}}[\log p(\mathbf{X}, \mathbf{Y}|\theta)] \tag{34}$$

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

$$\theta^{(i+1)} = \arg\max_\theta Q(\theta|\theta^{(i)}). \tag{35}$$

We solve for 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:

$$\frac{d}{d\theta[j]}Q(\theta|\theta^{(i)}) = \sum_p \sum_q \frac{dQ(\theta|\theta^{(i)})}{d\boldsymbol{A}[p,q]}\frac{d\boldsymbol{A}[p,q]}{d\theta[j]} \tag{36}$$

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

$$\frac{d\boldsymbol{A}}{d\theta[j]} = -i2\pi\theta[j]\boldsymbol{A}\left(\boldsymbol{u}\mathcal{M}_\varrho[:, j+1]^T + \boldsymbol{v}\mathcal{M}_\delta[:, j+1]^T\right). \tag{37}$$

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

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:

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

By expanding and taking the derivative of the log-likelihood from Equation 27 with respect to $\boldsymbol{A}$, we obtain the expression

$$\frac{d}{d\boldsymbol{A}} Q(\theta|\theta^{(i)}) = -\frac{1}{2}\sum_{t=2}^{N} \left[2\boldsymbol{Q}^{-1}\boldsymbol{A}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}^T\right] \right. \quad (38)$$
$$\left. -\boldsymbol{Q}^{-1}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_{t-1}^T\right] - \boldsymbol{Q}^{-1}E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right] \right].$$

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 $p(\boldsymbol{X}|\boldsymbol{Y};\theta^{(i)})$. Conveniently, these sufficient statistics can be computed using from the set of $\mathbf{z}$'s and $\mathbf{P}$'s computed in Section VI-A. From the marginal distributions (Equations 32 and 33) derived in Section VI-A, we obtain

$$E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_t\boldsymbol{x}_t^T\right] = \hat{\boldsymbol{x}}_t\hat{\boldsymbol{x}}_t^T + \mathbf{C}_t. \quad (39)$$

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 the supplemental material. Mathematically,

$$E_{\boldsymbol{X}|\boldsymbol{Y},\theta^{(i)}}\left[\boldsymbol{x}_{t-1}\boldsymbol{x}_t^T\right] = \hat{\boldsymbol{x}}_{t-1}\hat{\boldsymbol{x}}_t^T + \boldsymbol{\xi}_3\boldsymbol{\xi}_1^{T-1}, \quad (40)$$

where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_3$ are defined according to

$$p(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|\boldsymbol{Y}) = \mathcal{N}_{\boldsymbol{x}_{t-1}}(\boldsymbol{\xi}_1\boldsymbol{x}_t + \boldsymbol{\xi}_2, \boldsymbol{\xi}_3) \quad (41)$$
$$\propto \mathcal{N}_{\boldsymbol{x}_{t-1}}(\mathbf{z}_{t-1|t-1}^\alpha, \mathbf{P}_{t-1|t-1}^\alpha)\mathcal{N}_{\boldsymbol{x}_t}(\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{Q})\mathcal{N}_{\boldsymbol{x}_t}(\mathbf{z}_{t|t}^\beta, \mathbf{P}_{t|t}^\beta).$$

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 convergence has been reached, we return the parameters $\hat{\theta}$ and the optimal set of instantaneous images under that transformation $E[\boldsymbol{X}] = \{\hat{\boldsymbol{x}}_t\}_{t=1}^N$.

Note that since our EM method's maximization step requires solving a non-convex problem we likely will only find a local-maximum of θ at each step. Nonetheless, the log-likelihood is guaranteed to increase for a θ that increases $Q(\theta|\theta^{(i)})$ (Equation 34) [45]. This class of algorithms, which do not necessarily find the optimal θ at each iteration, are more rigorously referred to as "Generalized EM" [45]. In the case of a linear $f(\boldsymbol{x})$ this EM procedure is exact and the log-likelihood increases at every iteration. In the case of a non-linear $f(\boldsymbol{x})$ the forward-backward algorithm in Section VI-A provides only an approximation of the true sufficient statistics. Nonetheless, we empirically find that, when we fix each latent image's linearization point, the log-likelihood consistently improves.

VII. DYNAMIC IMAGING RESULTS

As data from the EHT 2017 campaign is yet to be released, in this section we demonstrate our method on synthetic EHT data and real data from the Very Long Baseline Array (VLBA). Additional results can be seen in the supplemental document and video. The StarWarps algorithm has been implemented as part of the publicly available python `eht-imaging`[5] library [6].

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

A. Synthetic Data Generation

We demonstrate our algorithm on synthetic data generated from four different sequences of time-varying sources. These sequences include two realistic fluid simulations of a black hole accretion disk for different observing orientations [48], a realistic sequence of a "hot spot" rotating around a black hole [47], 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 6. 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.

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 uv-coverage for this array can be seen in Figure 4. The second array, EHT2017+, augments the EHT2017 array with 3 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 VII-A. The third array, FUTURE, consists of 9 additional telescopes. The uv-coverage of these latter two arrays, along with a colorbar indicating the time of each measurement, is shown in Figures 7.

Visibility measurements are generated using the python `eht-imaging` library [6]. 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

$$\sigma = \frac{1}{0.88}\sqrt{\frac{\text{SEFD}_1 \times \text{SEFD}_2}{2 \times \Delta\nu \times t_{\text{int}}}}, \quad (42)$$

for System Equivalent Flux Density (SEFD) of the two telescopes corresponding to each visibility[6] [3]. Random station-based atmospheric phases drawn from a uniform distribution at each time step are 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.

B. 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.

1) Synthetic Data Result Comparison: Figure 8 shows example reconstructions, and corresponding measured error

[6]The factor of 1/0.88 is due to information loss due to recording 2-bit quantized data-streams at each telescope [1].
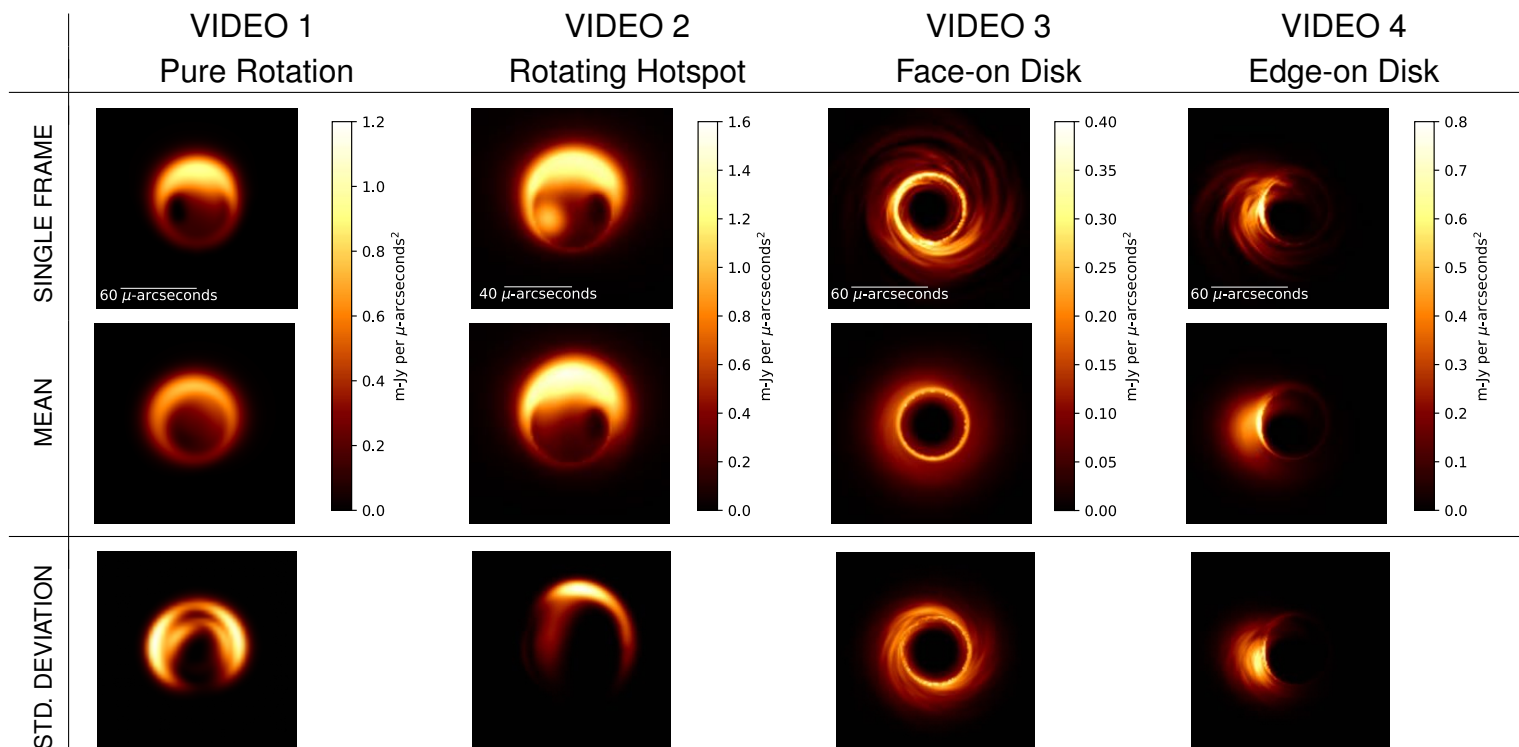
Fig. 6. **Ground truth videos:** The four ground truth sequences used to demonstrate results. We show a single frame from each sequence, the mean frame, and the spatial standard deviation of flux density. Video 1 consists of a 160μ-arcsecond image [46] that rotates $180°$ over the course of a 12 hour 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 [47] with a rotational period of 2.78 hours. Video 3 and 4 are generated using a model of a black hole observed face on and at a $45°$ inclination with a 160μ-arcsecond field of view [48]. They assume a spin of 0.9375 with an Innermost Stable Circular Orbit (ISCO) rotational period of 8.96 minutes. The specified FOV and colormaps for the single frame and mean images are used for each corresponding video throughout the remainder of the paper.
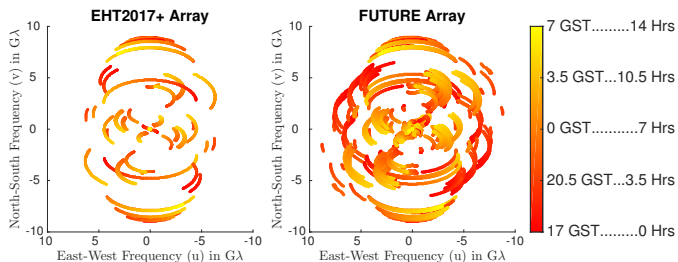
Fig. 7. **Time-varying uv-coverage:** The uv-coverage for EHT2017+ and FUTURE arrays when observing SgrA. The uv-coverage for the EHT2017 array can be seen in Figure 4. Baselines are colored by the time of each observation relative to the start time, indicated by the colorbar to the right.

TABLE I
EHT 2017 STATION PARAMETERS

Telescope	Location	SEFD (Jy)
ALMA	Chile	110
APEX	Chile	22000
LMT	Mexico	560
SMT	Arizona	11900
SMA	Hawaii	4900
JCMT	Hawaii	4700
PV	Spain	2900
SPT	South Pole	1600
PDB*	France	1600
HAY*	Massachusetts	2500
KP*	Arizona	2500

The location and SEFD of each telescope in the EHT2017 and EHT2017+ arrays. These parameters and locations were used to generate the uv-trajectories in frequency space shown in Figure 4 and 7. Telescope names followed by a star (*) were not included in the EHT2017 array.

(NRMSE), for combinations of the 4 source videos observed under the 3 telescope arrays. For these results we have set $a = 2$, $c = 1/2$, and $Q = 10^{-7}\mathbb{1}$. 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 6, but are smoothed to 3/4 the nominal resolution of the interferometer to help illustrate the level of resolution we aim to recover.

Horizontally within each quadrant 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
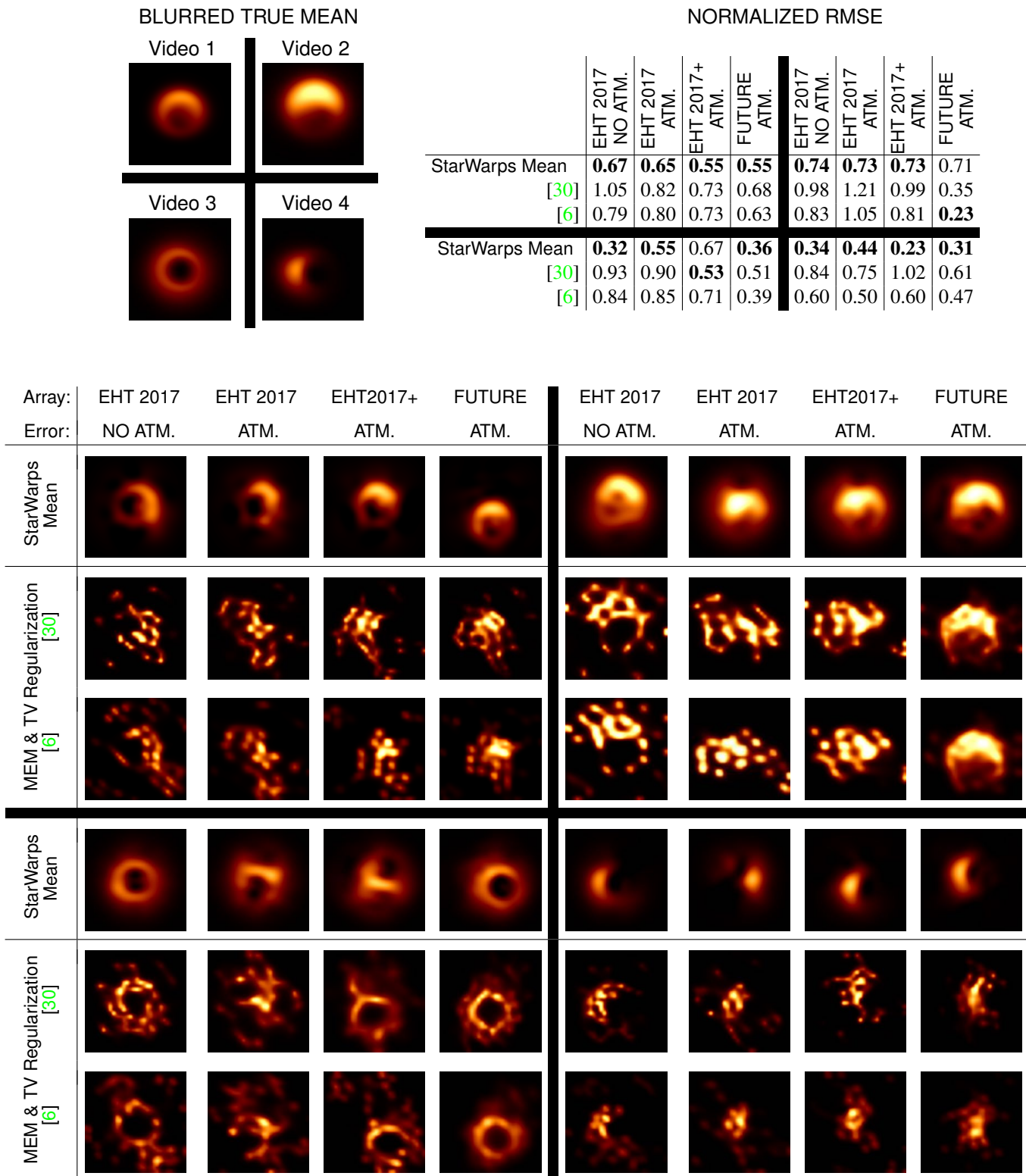
BLURRED TRUE MEAN

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
[30]	1.05	0.82	0.73	0.68	0.98	1.21	0.99	0.35
[6]	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**
[30]	0.93	0.90	**0.53**	0.51	0.84	0.75	1.02	0.61
[6]	0.84	0.85	0.71	0.39	0.60	0.50	0.60	0.47

Fig. 8. **Static evolution model:** Results obtained using data simulated from each of the 4 video sequences (see Figure 6) under different telescope arrays (see Figures 4 and 7) 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 [30] and [6]. 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. The FOV and colorbar used for each reconstruction can be seen in Figure 6.
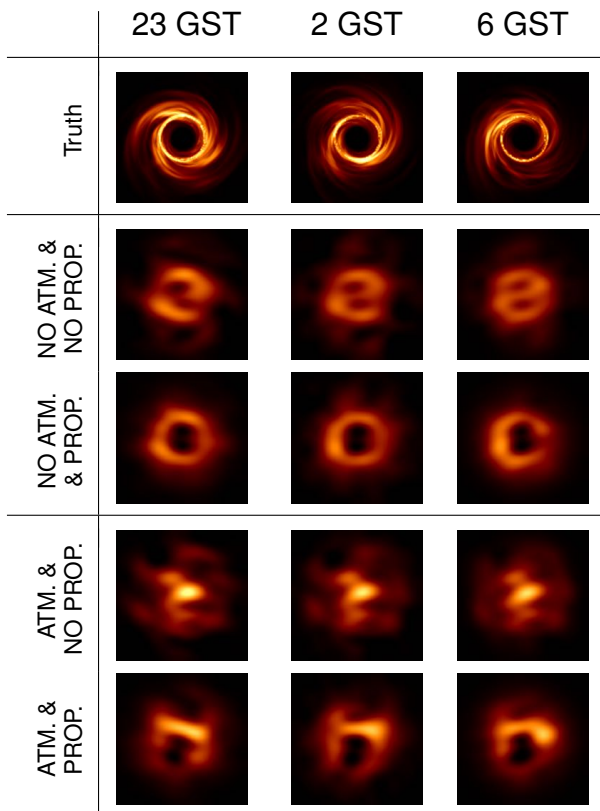
Fig. 9. **Propagating Uncertainty:** During inference, StarWarps approximates each image's covariance matrix in order to propagate its uncertainty to neighboring frames in time. Propagating this information is crucial when using very few measurements. We show frames resulting from the same EHT2017 data of Video 3 when the covariance is propagated (PROP.) as described in this paper, versus not propagated (NO PROP.). Note that propagating the covariance results in significantly improved results. This is true even in the case of using non-linear measurements when atmospheric error is present (ATM.). In this non-linear case the covariance matrix is simply a crude approximation of the uncertainty, but proves critical in obtaining a result that captures the ring structure of the underlying source.

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. [6] 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), however, in the case of an evolving source it often results in artifact-heavy reconstructions that are difficult to interpret. In [30] the authors attempt to mitigate 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 [30] are reconstructed by an author of the method.

2) The Importance in Propagating Uncertainty: StarWarps uses a multivariate Gaussian regularizer for imaging, which leads to a straightforward optimization method that propagates information through time. The uncertainty of each reconstructed image is encompassed in its approximated covariance matrix ($\mathbf{P}_{t|t}$), which informs the reconstruction of each neighboring latent image. Although this covariance matrix is sometimes a crude estimate of the true uncertainty, it is still
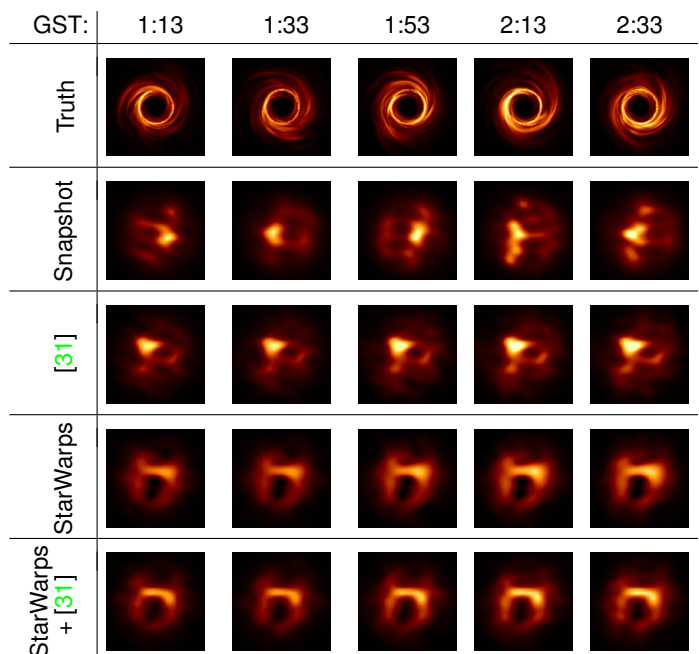

Fig. 10. **Comparison to [31] and Snapshot Imaging:** A comparison of results obtained using the proposed StarWarps method to Snapshot imaging and a method presented in [31]. The same simulated EHT2017 data of Video 3 was used for each result, and contained atmospheric noise. Snapshot imaging, which independently reconstructs each frame, is unable to produce reasonable results, and has no continuity through time due to the loss of absolute location information when using atmosphere corrupted measurements. The more flexible framework of [31] often makes it possible to obtain sharper and cleaner images, however struggles when working with very few measurements, as is the case for the EHT2017 array. Although results are consistent through time, [31] fails to recover the true ring structure of the source. StarWarps is able to begin recovering this ring structure, but contains a number of artifacts spurring from the main ring structure. Initializing [31] with the result of StarWarps produces a cleaner and sharper result.
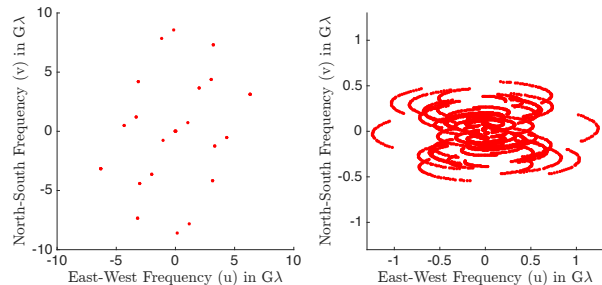
(a) Fig. 10 frame uv-coverage   (b) Fig. 12 frame uv-coverage

Fig. 11. **Single frame uv-coverage:** (a) The uv-coverage for the first frame shown in Figure 10 contains 21 measurements while (b) the uv-coverage for the first frame shown in Figure 12 contains 1736. When the measurements provided are very sparse, as in (a), StarWarps significantly outperforms [31]. However, in the case of many measurements, as in (b), [31] achieves better results with a higher dynamic range.

crucial in reconstructing faithful images when measurements are especially sparse.

The importance of propagating uncertainty through the covariance matrix is demonstrated in Figure 9. This figure shows the effect of turning on and off the covariance propagation. Covariance propagation can be easily turned off by setting $\mathbf{P}_{t|t} = 0$ at each forward and backward update. Results in the figure are shown on simulated data from the EHT2017 array on Video 3, with and without atmospheric error. Note that in both cases, propagating the covariance matrix helps to substantially improve results. This is true even even in the case of atmospheric error, when the measurement function
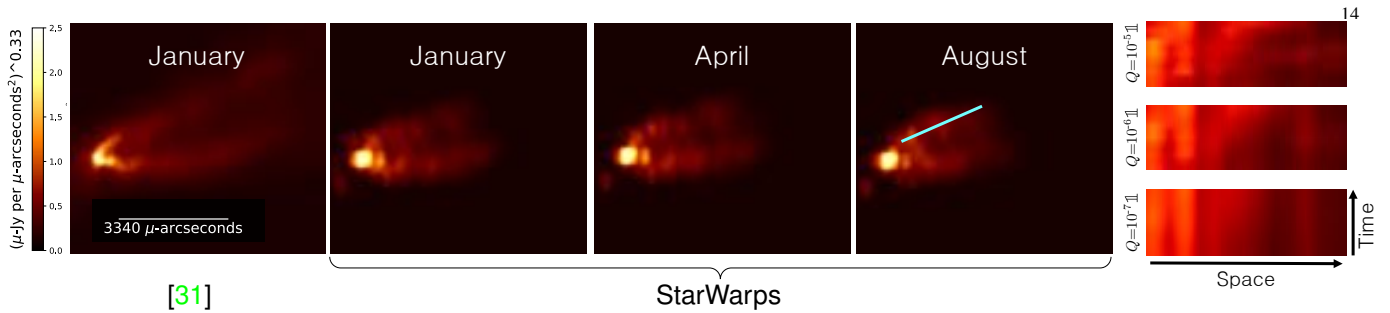
Fig. 12. **Video Reconstruction of Real Observations:** A StarWarps movie reconstruction obtained using real VLBI data taken of the M87 jet over the course of a year. Frames are shown with a gamma correction of $\gamma = 1/3$ to highlight weak emission. Data was collected in 2007 using the Very Long Baseline Array (VLBA) at 43 GHz [49]. As the source structure does not evolve over the course of a night, traditional imaging approaches can be used to reconstruct 'snapshot' images from this data. The forked structure appearing in the StarWarps reconstructions also appears in images reconstructed using the CLEAN static imaging approach [49] and the dynamic imaging approach presented in [31] (see left-most image). StarWarps produces a video that allows us to easily visualize the moving arms of the jet; the reconstructed video appears to contain outward motion, with a brighter region propagating down the arms. By visualizing the same slice of each frame (indicated by the cyan line) it becomes easier to see this motion as a static image (see the 3 Space × Time images on the far right). Note the diagonal 'line' shown in the top 2 of these Space × Time images indicates a bright region moves down the arm, towards the right of the image. The intensity of these slices has been increased by 70% to highlight the evolving, weak emission. By controlling the amount of temporal regularization, $\mathbf{Q}$, we control the amount of motion that appears in the reconstructed video. Increasing the temporal regularization, by decreasing $\mathbf{Q}$, results in a Space × Time slice that varies less with Time. Individual frames shown were generated using $\mathbf{Q} = 10^{-6}\mathbb{1}$.

$f(\mathbf{x})$ is non-linear and the covariance matrix is only a rough approximation of the true uncertainty.

3) Dynamical Imaging Comparison: As discussed in Section III, the Dynamical Imaging method presented in [31] was developed simultaneously, and shares many similarities to the work presented in this paper: they both aim to solve for a video rather than a static image. However, they have significant differences, leading to different strengths and weaknesses. The framework of [31] allows for more sophisticated image and temporal regularization, at the cost of a difficult optimization problem that does not propagate uncertainty. This results in sharper and cleaner videos when there is sufficient data, but can lead to poor results when there are very few measurements. Conversely, StarWarps' use of very simple Gaussian image and temporal regularization results in blurry results, but allows us to propagate an approximation of uncertainty (through the covariance matrix) and produce better results when very few measurements are available.

A comparison of results from [31] and StarWarps on EHT2017 simulated data can be seen in Figure 10. Results of [31] were produced using $\mathcal{R}_{\Delta I}$ and KL $\mathcal{R}_{\Delta t}$ temporal regularization, and Maximum Entropy and Total Variation Squared image regularization. Note that for this especially sparse data, [31] on its own does not faithfully reconstruct the ring structure of the underlying source. StarWarps is able to produce a ring, but with a number of blurry artifacts. Initializing [31] with the output of StarWarps produces the cleanest result. Although the StarWarps method runs faster on this example than [31] (84 seconds vs 204 seconds in Python on a 2.8 GHz Intel Core i7), StarWarps is memory intensive and its computational complexity scales poorly with increasing image size compared to [31]. To help solve these issues, in the future ideas from Ensemble Kalman Filters could be adapted in order to avoid StarWarp's costly matrix inversions and reduce the method's memory footprint [50].

An additional result comparing the two methods can be seen in Figure 12, which is discussed in the next section. In this example there is sufficient data to reconstruct each frame independently, and [31] is able to produce a cleaner image with a higher dynamic range than StarWarps.

Figure 11 compares the uv-coverage of a single frame for Figures 10 and Figure 12, highlighting that StarWarps is comparatively strongest in the case of sparse data, as will be available for the EHT. These examples demonstrate that StarWarps and [31] are complementary methods, and may ultimately lead to hybrid approaches for video reconstruction that produce higher quality results.

4) Application to Real VLBI Data: Although StarWarps was developed with the considerations of the EHT in mind, it can be applied to VLBI data taken from other sources and telescope arrays. For instance, galactic relativistic jet sources ("microquasars") often show variability over the course of a single observation [51]. However, due to physical constraints, most VLBI telescope networks observe sources that do not evolve this quickly, such as distant jets from the cores of Active Galactic Nuclei. In these cases, traditional static imaging approaches can be applied to each night of data to produce faithful reconstructions. Yet, by jointly processing the data taken over a larger span of time, we are able to make movies of long-term source evolution that preserve continuity of features through time, thus reducing the flickering that occurs when independently reconstructing each frame.

In Figure 12 we demonstrate StarWarps on archival data taken of the M87 jet. This data was taken using the Very Long Baseline Array (VLBA) as part of the M87 Movie Project [49]. Ten epochs of data between the beginning of January and end of August in 2007 were processed simultaneously. Images were reconstructed with a 10 m-arcsecond field of view with $M = 70$ pixels.

Unlike as expected in EHT observations, the dynamic range of the M87 Jet is very high. In order to faithfully reconstruct a high dynamic range image using the simple Gaussian prior, we have incorporated gamma correction into our measurement function. Rather than reconstruct a video containing linear-scale images, we instead reconstruct gamma-corrected images. To do this we replace the measurement function $f(\mathbf{x})$ with $f(\mathbf{x}^{\frac{1}{\gamma}})$. During reconstruction of this M87 Jet video we have used $\gamma = 1/2$. Although these images still do not have the same dynamic range that is achieved through other imaging methods [49], [31], StarWarps is still able to recover the faint arms of the jet.

The reconstructed movie produced by StarWarps shows

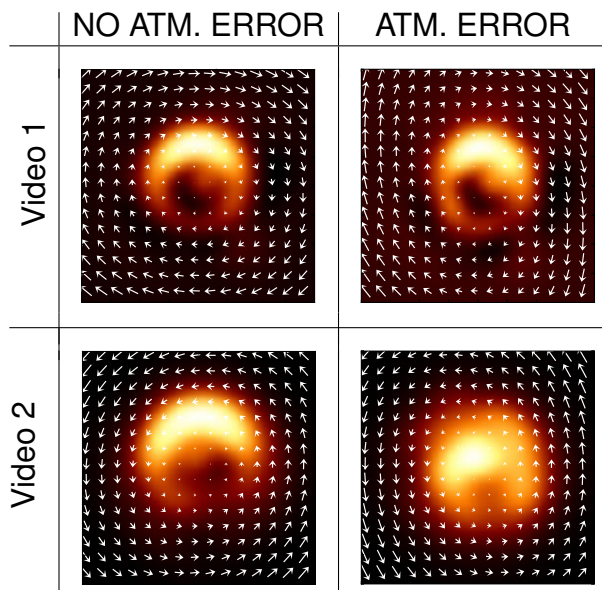**NO ATM. ERROR** | **ATM. ERROR**

Video 1

Video 2

Fig. 13. **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 (ATM. and NO ATM. ERROR, respectively). Arrows showing the direction of recovered motion are overlaid on the mean image for a recovered video. Refer to the supplemental video for a visualization of the true underlying and recovered videos. 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.

outward motion along the jet. While this motion is hard to see in Figure 12's static frames, by visualizing a slice of each image (indicated by the cyan line) through time the motion becomes more apparent. The resulting Space × Time image shows a brighter region of emission moving along the arm of the jet. We show the same Space × Time reconstruction for different weightings of temporal regularization, Q. Note that as temporal regularization increases, by decreasing Q, the Space × Time image becomes more uniform in time.

C. Unknown Evolution Model (Learn Warp)

In Section VII-B 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 VI-B, 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 13 shows the 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), there does not exist a persistent warp field that fits the data, let alone an affine warp field. Although the true motion

Single Frame with Overlaid Circle | Unwrapped Space × Time Image

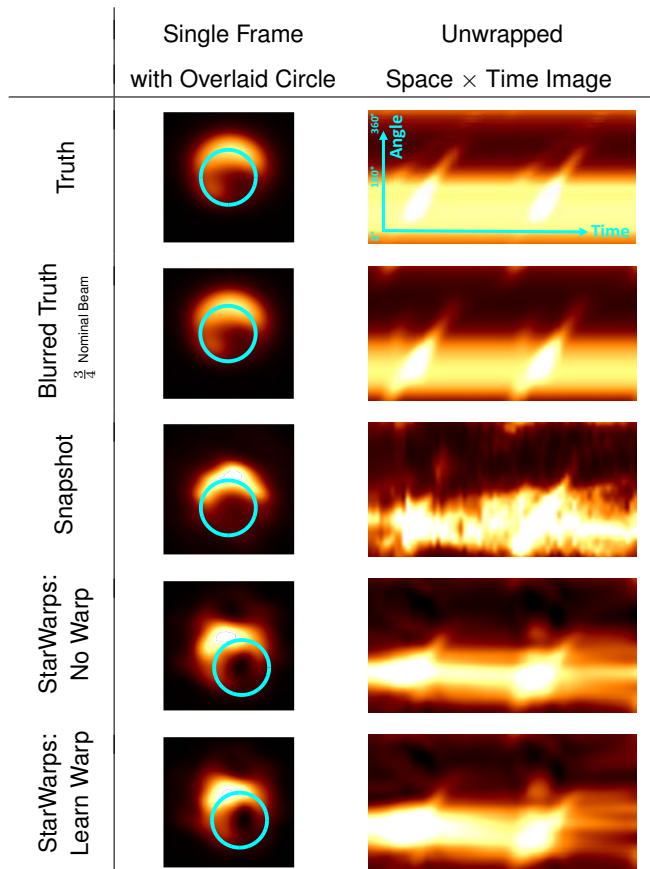Truth — Blurred Truth — Snapshot — StarWarps: No Warp — StarWarps: Learn Warp

Fig. 14. **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 2 of the supplemental document. 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.

cannot be described by our model, we still recover an accurate estimate indicating the direction of motion.

Figure 14 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 is independently reconstructed using only the small number of measurements taken at that time step. In particular, we use the MEM & TV method shown in Figure 4 to reconstruct each snapshot. Our results using StarWarps show substantial improvement over snapshot imaging, especially in the case of data containing atmospheric phase error.

We expand upon these results in the supplemental material's video and document. Figures 1 and 2 in the supplemental material document 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. 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.

VIII. Conclusion

Traditional interferometric imaging methods are designed under the assumption that the target source is static over the course of an observation [1]. 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 [11]. 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.

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.

References

[1] Thompson, A.R., Moran, J.M., Swenson, Jr., G.W.: Interferometry and Synthesis in Radio Astronomy, 3rd Edition. (2017) 1, 2, 3, 5, 10, 16
[2] Högbom, J.: Aperture synthesis with a non-regular distribution of interferometer baselines. Astron. Astrophys. Suppl (1974) 1, 4
[3] Taylor, G.B., et. al: Synthesis imaging in radio astronomy ii. In: Synthesis Imaging in Radio Astronomy II. (1999) 1, 3, 4, 7, 10
[4] Lu, R.S., Broderick, A.E., Baron, F., Monnier, J.D., Fish, e.a.: Imaging the Supermassive Black Hole Shadow and Jet Base of M87 with the Event Horizon Telescope. The Astrophysical Journal (June 2014) 1, 5
[5] Bouman, K.L., Johnson, M.D., Zoran, D., Fish, V.L., Doeleman, S.S., Freeman, W.T.: Computational imaging for vlbi image reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 913–922 1, 3, 4, 5, 6
[6] Chael, A.A., Johnson, M.D., Narayan, R., Doeleman, S.S., Wardle, J.F.C., Bouman, K.L.: High-resolution linear polarimetric imaging for the event horizon telescope. The Astrophysical Journal 829(1) (2016) 11 1, 4, 5, 6, 10, 12, 13
[7] Akiyama, K., Kuramochi, K., Ikeda, S., Fish, V.L., Tazaki, F., Honma, M., Doeleman, S.S., Broderick, A.E., Dexter, J., Mocibrodzka, M., Bouman, K.L., Chael, A.A., Zaizen, M.: Imaging the schwarzschild-radius-scale structure of m87 with the event horizon telescope using sparse modeling. The Astrophysical Journal 838(1) (2017) 1 1, 5
[8] Fish, V., Akiyama, K., Bouman, K., Chael, A., Johnson, M.e.a.: Observing—and Imaging—Active Galactic Nuclei with the Event Horizon Telescope. Galaxies 4 (October 2016) 54 1, 4
[9] Akiyama, K., Ikeda, S., Pleau, M., Fish, V.L., Tazaki, F.e.a.: Super-resolution full-polarimetric imaging for radio interferometry with sparse modeling. The Astronomical Journal 153(4) (2017) 159 1
[10] Ghez, A.M., Salim, S., Weinberg, N.N., Lu, J.R., Do, T.e.a.: Measuring Distance and Properties of the Milky Way's Central Supermassive Black Hole with Stellar Orbits. ApJ689 (December 2008) 1044–1062 1
[11] Johnson, M.D., Loeb, A., et. al, H.S.: Measuring the direction and angular velocity of a black hole accretion disk via lagged interferometric covariance. The Astrophysical Journal 813(2) (2015) 132 1, 16
[12] Fish, V.L., et. al, S.S.D.: 1.3 mm wavelength vlbi of sagittarius a*: Detection of time-variable emission on event horizon scales. The Astrophysical Journal Letters 727(2) (2011) L36 1
[13] Johnson, M.D., Fish, V.L., Doeleman, S.S., Marrone, D.P., Plambeck, R.L.e.a.: Resolved magnetic-field structure and variability near the event horizon of Sagittarius A*. Science 350 (December 2015) 1242–1245 1
[14] Doeleman, S., Agol, E., Backer, D., Baganoff, F., Bower, G.C.e.a.: 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) 1
[15] Doeleman, S.S., Weintroub, J., Rogers, A.E.E.e.a.: Event-horizon-scale structure in the supermassive black hole candidate at the Galactic Centre. Nature455 (September 2008) 78–80 1, 2
[16] Johannsen, T., Psaltis, D.: Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. II. Black Hole Images. ApJ718 (July 2010) 446–454 1
[17] Monnier, J.D., Allen, R.J.: Radio and optical interferometry: Basic observing techniques and data analysis. In: Planets, Stars and Stellar Systems. Springer (2013) 3
[18] Jennison, R.C.: A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. MNRAS118 (1958) 276 3
[19] Bracewell, R.: Fourier analysis and imaging. Springer Science & Business Media (2004) 3

[20] Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic resonance in medicine 58(6) (2007) 1182–1195 3
[21] Munson, D.C., O'Brien, J.D., Jenkins, W.K.: A tomographic formulation of spotlight-mode synthetic aperture radar. Proceedings of the IEEE (Aug 1983) 3
[22] Demanet, L., Jugnon, V.: Convex recovery from interferometric measurements. IEEE Transactions on Computational Imaging 3(2) (2017) 282–295 3
[23] Narayan, R., Nityananda, R.: Maximum entropy image restoration in astronomy. ARA&A24 (1986) 127–170 4
[24] Skilling, J.: Quantified maximum entropy. In: Maximum Entropy and Bayesian Methods. Springer (1990) 4
[25] Baron, F., Monnier, J.D., Kloppenborg, B.: A novel image reconstruction software for optical/infrared interferometry. In: SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics (2010) 77342I–77342I 4
[26] Buscher, D.: Direct maximum-entropy image reconstruction from the bispectrum. In: Very High Angular Resolution Imaging. Springer (1994) 91–93 4, 5
[27] Suksmono, A.B.: Deconvolution of vlbi images based on compressive sensing. In: 2009 International Conference on Electrical Engineering and Informatics. Volume 01. (Aug 2009) 110–116 4
[28] Girard, J., Garsden, H., Starck, J., Corbel, S., Woiselle, A., Tasse, C., McKean, J., Bobin, J.: Sparse representations and convex optimization as tools for lofar radio interferometric imaging. Journal of Instrumentation 10(08) (2015) C08013 4
[29] Doeleman, S.S., Fish, V.L., Broderick, A.E., Loeb, A., Rogers, A.E.E.: Detecting flaring structures in sagittarius a* with high-frequency vlbi. The Astrophysical Journal 695(1) (2009) 59 4
[30] Lu, R.S., Roelofs, F., Fish, V.L., Shiokawa, H., Doeleman, S.S., Gammie, C.F., Falcke, H., Krichbaum, T.P., Zensus, J.A.: Imaging an event horizon: Mitigation of source variability of sagittarius a*. The Astrophysical Journal 817(2) (2016) 173 4, 12, 13
[31] Johnson, M.D., Bouman, K.L., Blackburn, L., Chael A. A, R.J.e.a.: Dynamical Imaging With Interferometry. The Astrophysical Journal 850(1) (December 2017) 172 4, 13, 14
[32] Anderson, B.D., Moore, J.B.: Optimal filtering. Englewood Cliffs 21 (1979) 22–95 4, 9
[33] Tsao, J., Kozerke, S.: Mri temporal acceleration techniques. Journal of Magnetic Resonance Imaging 36(3) (2012) 543–560 4
[34] Liang, Z.P., Lauterbur, P.C.: An efficient method for dynamic magnetic resonance imaging. IEEE Transactions on medical imaging 13(4) (1994) 677–686 4
[35] Khalsa, K.A., Fessler, J.A.: Resolution properties in regularized dynamic mri reconstruction. In: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. (April 2007) 456–459 4, 8
[36] Chun, S.Y., Fessler, J.A.: Spatial resolution properties of motion-compensated tomographic image reconstruction methods. IEEE transactions on medical imaging 31(7) (2012) 1413–1425 4
[37] Nakarmi, U., Wang, Y., Lyu, J., Liang, D., Ying, L.: A kernel-based low-rank (klr) model for low-dimensional manifold recovery in highly accelerated dynamic mri. IEEE transactions on medical imaging 36(11) (2017) 2297–2307 4
[38] Chen, G.H., Tang, J., Leng, S.: Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. Medical physics 35(2) (2008) 660–663 4
[39] Biswas, S., Dasgupta, S., Mudumbai, R., Jacob, M.: Subspace aware recovery of low rank and jointly sparse signals. IEEE Transactions on Computational Imaging 3(1) (2017) 22–35 4
[40] Torralba, A., Oliva, A.: Statistics of natural image categories. Network: computation in neural systems 14(3) (2003) 391–412 5
[41] Hammersley, J.M., Clifford, P.: Markov fields on finite graphs and lattices. (1971) 7
[42] Mobahi, H., Liu, C., Freeman, W.T.: A compositional model for low-dimensional image set representation. In: 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014), IEEE Computer Society (2014) 8
[43] Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(2) (Feb 2006) 236–250 8
[44] Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning. The MIT Press (2009) 8, 9
[45] Little, R.J., Rubin, D.B.: Statistical analysis with missing data. John Wiley & Sons (2014) 10
[46] Broderick, A.E., Fish, V.L., Doeleman, S.S., Loeb, A.: Evidence for Low Black Hole Spin and Physically Motivated Accretion Models from Millimeter-VLBI Observations of Sagittarius A*. The Astrophysical Journal (July 2011) 11
[47] Broderick, A.E., Loeb, A.: Imaging optically-thin hotspots near the black hole horizon of Sgr A* at radio and near-infrared wavelengths. MNRAS367 (April 2006) 905–916 10, 11
[48] Shiokawa, H.: General-relativistic magnetohydrodynamics simulations of black hole accretion disks: Dynamics and radiative properties. PhD thesis, University of Illinois at Urbana-Champaign (2013) 10, 11
[49] Walker, R.C., Hardee, P.E., Davies, F., Ly, C., Junor, W., Mertens, F., Lobanov, A.: Observations of the structure and dynamics of the inner m87 jet. Galaxies 4(4) (2016) 46 14
[50] Evensen, G.: The ensemble kalman filter: Theoretical formulation and practical implementation. Ocean dynamics 53(4) (2003) 343–367 14
[51] Frazin, R.A., Butala, M.D., Kemball, A., Kamalabadi, F.: Time-dependent reconstruction of nonstationary objects with tomographic or interferometric measurements. The Astrophysical Journal Letters 635(2) (2005) L197 14