Deep learning inference with the Event Horizon Telescope

I. Calibration improvements and a comprehensive synthetic data library

M. Janssen[1,2], C.-k. Chan[3,4,5], J. Davelaar[6,7], I. Natarajan[8,9], H. Olivares[10], B. Ripperda[11,12,13,14], J. Röder[2], M. Rynge[15], and M. Wielgus[16]

1. Department of Astrophysics, Institute for Mathematics, Astrophysics and Particle Physics (IMAPP), Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
   e-mail: M.Janssen@astro.ru.nl
2. Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany
3. Steward Observatory and Department of Astronomy, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA
4. Data Science Institute, University of Arizona, 1230 N. Cherry Ave., Tucson, AZ 85721, USA
5. Program in Applied Mathematics, University of Arizona, 617 N. Santa Rita Ave., Tucson, AZ 85721, USA
6. Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
7. NASA Hubble Fellowship Program, Einstein Fellow
8. Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
9. Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA
10. Departamento de Matemática da Universidade de Aveiro and Centre for Research and Development in Mathematics and Applications (CIDMA), Campus de Santiago, 3810-193 Aveiro, Portugal
11. Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, ON, Canada M5S 3H8
12. David A. Dunlap Department of Astronomy, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4
13. Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON M5S 1A7
14. Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
15. University of Southern California - Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292, USA
16. Instituto de Astrofísica de Andalucía-CSIC, Glorieta de la Astronomía s/n, E-18008 Granada, Spain

ABSTRACT

Context. In a series of publications, we describe a comprehensive comparison of Event Horizon Telescope (EHT) data with theoretical models of the observed Sagittarius A* (Sgr A*) and Messier 87* (M87*) horizon-scale sources.

Aims. In this first article, we report on improvements made to our observational data reduction pipeline and present the generation of observables derived from the EHT models. We make use of ray-traced general relativistic magnetohydrodynamic simulations that are based on different black hole spacetime metrics and accretion physics parameters. These broad classes of models provide a good representation of the primary targets observed by the EHT.

Methods. We describe how we combined multiple frequency bands and polarization channels of the observational data to improve our fringe-finding sensitivity and stabilization of atmospheric phase fluctuations. To generate realistic synthetic data from our models, we took the signal path as well as the calibration process, and thereby the aforementioned improvements, into account. We could thus produce synthetic visibilities akin to calibrated EHT data and identify salient features for the discrimination of model parameters.

Results. We have produced a library consisting of an unparalleled 962,000 synthetic Sgr A* and M87* datasets. In terms of baseline coverage and noise properties, the library encompasses 2017 EHT measurements as well as future observations with an extended telescope array.

Conclusions. We differentiate between robust visibility data products related to model features and data products that are strongly affected by data corruption effects. Parameter inference is mostly limited by intrinsic model variability, which highlights the importance of long-term monitoring observations with the EHT. In later papers in this series, we will show how a Bayesian neural network trained on our synthetic data is capable of dealing with the model variability and extracting physical parameters from EHT observations. With our calibration improvements, our newly reduced EHT datasets have a considerably better quality compared to previously analyzed data.

Key words. accretion, accretion disks – black hole physics – techniques: high angular resolution – techniques: interferometric – galaxies: active

1. Introduction

Powered by accretion onto supermassive black holes (SMBHs), active galactic nuclei (AGNs) belong to the most luminous persistent sources in the known Universe (Schmidt 1963; Lynden-Bell 1969). With the Event Horizon Telescope (EHT) very long baseline interferometry (VLBI) array, we can resolve the innermost region of AGNs and study accretion onto compact objects, plasma physics, jet launching, and gravity in the strong field regime (Event Horizon Telescope Collaboration et al. 2019a,b; Kim et al. 2020; Psaltis et al. 2020; Event Horizon Telescope Collabora-

tion et al. 2021; Kocherlakota et al. 2021; Janssen et al. 2021; Event Horizon Telescope Collaboration et al. 2022a,e; Issaoun et al. 2022; Jorstad et al. 2023; Paraschos et al. 2024; Baczko et al. 2024). However, interpreting EHT data poses significant challenges, such as the inherent variability of black hole accretion flows, atmospheric noise, and instrument-related data corruption effects. Addressing these issues requires both methodological advances and high-fidelity simulations of synthetic data.

Currently, general relativistic magnetohydrodynamics (GRMHD) simulations (e.g., Mizuno 2022) are the most complete class of models available for a self-consistent description of AGNs' central activity. GRMHDs naturally produce jets and, without (substantial) radiative cooling, a geometrically thick accretion flow. For parameters matching the Sagittarius A* (Sgr A*) and Messier 87* (M87*) sources, general relativistic ray-tracing (GRRT) calculations based on GRMHD simulations predict (sub-)millimeter synchrotron emission to be produced in an (mostly) optically thin plasma surrounding the black hole (Event Horizon Telescope Collaboration et al. 2019e, 2022f). From GRMHD-GRRT models, full Stokes horizon-scale movies of the disk plus jet emission of Sgr A* and M87* have been made. The geometrically thick and optically thin accretion scenario with accompanying jets is suitable for low-powered AGNs (e.g., Ho 2008, and references therein). The activity of luminous AGNs and X-ray binaries can also be simulated with GRMHD-GRRT models (e.g., Dexter et al. 2021; Liska et al. 2021; Wielgus et al. 2022a), but unlike low accretion rate sources, they generally require radiative cooling to be self-consistent. For the EHT, GRMHD-GRRT simulations are used for the validation of image reconstructions (Event Horizon Telescope Collaboration et al. 2019d, 2022c) and tests of general relativity (GR) together with the inference of MHD, accretion, and black hole parameters (Event Horizon Telescope Collaboration et al. 2019e,f; Psaltis et al. 2020; Kocherlakota et al. 2021; Event Horizon Telescope Collaboration et al. 2022f,e).

In this first paper of a series, we present a comprehensive synthetic data library that is based on a wide range of GRMHD-GRRT models. Specifically, we used three broad classes of models with different spacetime metrics. The first class is based on the Kerr solution to Einstein's field equations in GR without charge (Kerr 1963). The second class is based on the Kerr-Newman solution to GR, which describes a rotating, charged black hole (Newman et al. 1965). The third class is a dilaton black hole, which is a solution of the Einstein-Maxwell-dilaton-axion field equations (García et al. 1995). Beyond GR, the dilaton scalar field arises in the low-energy regime of heterotic string theory (Gross et al. 1985).

The forward modeling for the synthetic data generation follows the EHT-VLBI data path to make accurate predictions of what the instrument would measure given a known ground-truth sky brightness distribution. Multiple realizations of the synthetic datasets are produced to account for uncertain nuisance effects, such as telescope gain errors. The characteristics of the mock data should be as similar to the observational data as possible. In Janssen et al. (2025a), the second paper in this series, the synthetic data library is used to train a Bayesian neural network for GRMHD-GRRT parameter inference and present alongside a demonstration of how this inference can be improved with planned upgrades to the EHT. In the third paper (Janssen et al. 2025b), parameter posteriors are found by applying the trained neural network to current EHT observations.

This paper is organized as follows: in Section 2, we report on improvements made to our EHT data calibration methods, which are also used for synthetic data generation. In Section 3,
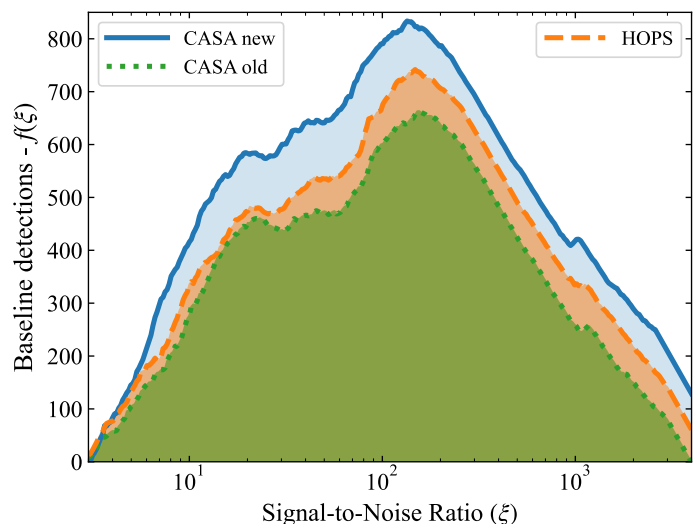
Fig. 1. Comparison of the accumulative number of detections of the 226.1 - 228.1 GHz EHT data of Sgr A* from 7 April 2017 and M87* data from 11 April 2017 from different data reductions. The visibilities are averaged into a single frequency channel and time-averaged into 120 s bins. The signal-to-noise ratio (ξ) is computed from the total intensity data (averaged parallel-hand correlation products after the polarization calibration). Only data with $\xi > 3$ are considered as detections, which are counted on all baselines. The detections are plotted as cumulative distributions minus the function $f(\xi) = 280 \log(\xi) - 305$.

we describe the GRMHD-GRRT image library that contains the ground-truth models of Sgr A* and M87* for our synthetic data generation of EHT observations. In Section 4, we list all effects taken into account for our forward modeling. In Section 5, we depict the data generation pipeline from an algorithmic point of view together with the content, format, and availability of the produced synthetic data library. In Section 6, we describe the salient features of our synthetic data, focusing on characteristics that allow us to discriminate between ground truth model parameters in the presence of the corruption effects that are present in EHT observations. We offer our conclusions about the GRMHD-GRRT synthetic data library for the EHT in Section 7.

2. Updated observational EHT data calibration

In this section, we present updates to our observational data reduction with respect to the M87* and Sgr A* calibration methods described in Event Horizon Telescope Collaboration et al. (2019c) and Event Horizon Telescope Collaboration et al. (2022b), respectively. For the forward modeling synthetic data generation, we also emulate the VLBI data calibration process. This process is based on the updated procedures described below. Furthermore, the parameter inference presented in Janssen et al. (2025b) is based on observational data reduced with these updated methods.

2.1. Signal stabilization

Recent upgrades to the Common Astronomy Software Applications (CASA) suite (CASA Team et al. 2022; van Bemmel et al. 2022) and enhancements to the Radboud PIpeline for the Calibration of high Angular Resolution Data (RPICARD) calibration pathway (Janssen et al. 2018, 2019b) have increased fringe detection sensitivity by 10% across all baselines, mainly at low signal-to-noise ratio (S/N), as shown in Figure 1. Previously, the signal stabilization through fringe-fitting and atmospheric

phase calibration (Janssen et al. 2022) was done individually for the parallel-hand correlation products and each of the EHT's 1875 MHz frequency bands. Now, after the instrumental phase and delay alignment steps, all four correlation products are combined, each over their full bandwidth, which is 3750 MHz in aggregate in 2017 and 7500 MHz in aggregate for the 2018 and subsequent EHT observations with the current instrumentation.

2.2. A priori flux density calibration

We fit the telescope gains and estimated their uncertainties for each polarization channel individually, as described in Janssen et al. (2019a) and Event Horizon Telescope Collaboration et al. (2019c, 2022b). Assuming the primary calibrator sources to have negligible intrinsic circular polarization, the dominant uncertainties from planet models and scatter in the antenna temperature (T_a^*) measurements (due to weather for example) are the same for both polarizations. From the relative T_a^* scatter between the polarizations, we estimate that the relative uncertainty is typically at the 1 % level, which is in agreement with the gain modeling presented in Event Horizon Telescope Collaboration et al. (2021). This relative uncertainty is used when drawing gain errors for synthetic data realizations to avoid an overestimation of the data corruption through telescope gains, which could otherwise induce an artificial circular polarization signal in the visibility amplitudes. Static, polarization-independent gain offsets are typically at the ~ 10 % level (Janssen et al. 2019a).

Moreover, we apply the system temperature measurements prior to the fringe fitting. Thereby, we took measured differences in the sensitivities within signal stabilization solution intervals into account. Additionally, we correctly applied frequency-resolved system temperatures to the corresponding frequency-resolved data. Previously, we have used custom post-processing scripts to apply the flux density calibration metadata to the per-band frequency-averaged data after the signal stabilization. In the case of the Atacama Large Millimeter/submillimeter Array (ALMA), which provides frequency-resolved calibration measurements at a 58 MHz level (Goddi et al. 2019), a geometric mean of the system temperatures had previously been applied to the visibilities.

2.3. Comparison with previous data products

The combined effect of the aforementioned improvements is visualized in Figure 1. For the Sgr A* data from 7 April 2017 and M87* data from 11 April 2017, we compare the data previously produced by Rpicard and eht-hops (Blackburn et al. 2019) as described in Event Horizon Telescope Collaboration et al. (2019c) and Event Horizon Telescope Collaboration et al. (2022b) with the improved new Rpicard data described here. The old Rpicard:casa data and corresponding old eht-hops:hops data were released under the 10.25739/g85n-f134[1] and 10.25739/m140-ct59[2] digital object identifiers. For Rpicard, the old data was obtained with v3.13.1 of the pipeline, which is also containerized with the `30e6ca14fb50275013c668285a3b476f9bc85436_91da63236db34f3a31b5309b18ac159128f28a35`[3] tag. The new data was obtained with Rpicard v7.2.2, containerized under `646d6a189c01b04cfa10077a46650038d61687d9_25c42c3c75a8334d1be4f72bc56b4344dc1f068e`[4].

---

[1] https://doi.org/10.25739/g85n-f134.
[2] https://doi.org/10.25739/m140-ct59.
[3] https://tinyurl.com/3p2dy5yw.
[4] https://tinyurl.com/2avph5rd.

The total number of fringe detections is determined by the sensitivity of a data reduction pathway. Imperfect corrections for instrumental phase and delay errors over a telescope bandpass or an incorrect estimation of the thermal noise for example could lead to missed or misidentified fringes. The distribution of the number of detections versus S/N is indicative of the data quality. Decoherence effects such as residual delay errors and uncorrected atmospheric phase turbulence can shift detections to lower S/N for the averaged visibilities.

From Figure 1, we observe that all data reduction pathways follow roughly the same distribution of detections. Compared to the old casa data, eht-hops has additional detections to Hawai'i telescopes, which were dropped in the global fringe search of old casa as the baselines could not be connected to the remaining telescopes of the EHT array. Additionally, the hops data contains additional detections in the 3 – 4 S/N range compared to the old casa data. These could either arise from a more conservative fringe rejection threshold used in casa or from the robust ad hoc phasing algorithm used in eht-hops to correct for atmospheric phase turbulence (Blackburn et al. 2019). This algorithm can establish sufficient phase coherence for additional detections at low S/N.

Compared to the old casa data, in the new data reduction presented here, sufficient S/N is accumulated over the combined EHT bands to connect all Hawai'i fringes to the rest of the array. No detections are lost at low S/N and excess detections over hops and the old casa are obtained at an S/N of around five. For the other two reduction pathways, the sensitivity and atmospheric phase stabilization are not sufficient to obtain these detections when integrating for 120 s. The new casa data has overall the most detections over the full S/N range of the data. Further upgrades of our EHT data processing will be enabled through the ngHOPS (Hoak et al. 2022) and ngCASA (van Bemmel et al. 2022; CASA Team et al. 2022) projects.

3. Simulation models

We generated synthetic datasets based on M87* and Sgr A* GRMHD-GRRT models. These datasets match the 2017 EHT observations of M87* (Event Horizon Telescope Collaboration et al. 2019c) and Sgr A* (Event Horizon Telescope Collaboration et al. 2022b) in terms of (u, v) coverage and S/N. Our "standard" parameter set is based on the fiducial kharma Kerr black hole EHT models from the patoka pipeline (Wong et al. 2022; Prather et al. 2021) that are used in Event Horizon Telescope Collaboration et al. (2019e, 2022f) and ray-traced in all Stokes parameters with ipole (Mościbrodzka & Gammie 2018). In Sections 3.1 to 3.7 below, we describe the broad range of parameters that are sampled by our standard models. Additional "exotic" models used for the synthetic data generation are introduced in Section 3.8. The underlying simulations are based on alternative black hole solutions beyond the Kerr metric and sample a smaller parameter space compared to the standard models. We note here the generally good agreement between the well-tested different GRMHD and GRRT codes used in the EHT collaboration (Porth et al. 2019; Gold et al. 2020; Prather et al. 2023).

3.1. Magnetic field configuration

Depending on the amount of magnetic flux Φ_{BH} crossing the black hole event horizon, different classes of accretion states develop (Event Horizon Telescope Collaboration et al. 2019e, 2022f). For high Φ_{BH}, magnetically arrested disk (MAD) states form, when the magnetic flux on the horizon exceeds a threshold

where the magnetic pressure becomes larger than the disk's ram pressure. During such a magnetic flux eruption, part of the accumulated magnetic energy is dissapated. Due to the large amount of magnetic flux, powerful jets are launched via the Blandford-Znajek mechanism if the black hole is also rotating (Blandford & Znajek 1977). For an accretion rate $\dot{M}$, gravitational radius r_g, and speed of light c, a MAD accretion state is reached when $\phi_{mag} \equiv \Phi_{BH} \left(\dot{M} r_g^2 c \right)^{-0.5} \gtrsim 50$ in Gaussian units. For $\phi_{mag} \sim 1$, a standard and normal evolution (SANE) state without flux eruptions develops.

The GRMHD simulations were initialized with a weakly magnetized Fishbone-Moncrief (Fishbone & Moncrief 1976) torus surrounding the black hole. MAD states were seeded by concentrating the initial magnetic field toward the inner edge of the torus. The torus is in hydrostatic equilibrium and small perturbations were added to the magnetic field. Instabilities, such as the magnetorotational instability (MRI, Balbus & Hawley 1998) develop, and the subsequent turbulence triggers accretion onto the black hole. Without cooling, this setup self-consistently develops a radiatively inefficient accretion flow (RIAF) plus jet state.

3.2. Spin

For a rotating Kerr black hole with mass M and angular momentum J, the dimensionless black hole spin is given as

$$a_* = \frac{Jc}{GM^2} , \tag{1}$$

where G is the gravitational constant. Our standard GRMHD simulations have $a_* = \pm 0.94, \pm 0.5, 0$. The sign of the angular momentum is given with respect to the accretion flow; negative spins correspond to counter-rotation.

3.3. Inclination angle

For Sgr A*, GRRT images were created for inclination angles of $i_{los} = 10°, 30°, 50°, 70°, 90°, 110°, 130°, 150°,$ and $170°$, with respect to the accretion flow angular momentum vector. Inclinations in the $(180, 360)$ degree range are related to the inclinations in the $(0, 180)$ degree range up to image plane rotation and stochastic effects.

For M87*, the inclination is known to be $17.2° \pm 3.3°$ based on cm VLBI observations (Mertens et al. 2016). Here, we assume that the jet seen on larger scales is aligned with the spin axis of the black hole. In the simulations, $i_{los} = 17°$ is used for $a_* < 0$ and $i_{los} = 163°$ for $a_* \geq 0$ following the Event Horizon Telescope Collaboration et al. (2019e).

3.4. Position angle

For Sgr A*, images were rotated by $\theta_{PA} = 0°, 30°, 60°, 90°, 120°, 150°,$ and $180°$ in the plane of the sky. This rotation was performed in the (u, v) plane during the synthetic data generation.

For M87*, the position angle was fixed to $288°$ based on the orientation of the large-scale jet (Walker et al. 2018). While the large scale jet is known to change orientation on the timescale of years (Walker et al. 2018; Cui et al. 2023), GRMHD model fitting to the 11 April 2017 EHT data of M87* found θ_{PA} values close to the average $288°$ (Event Horizon Telescope Collaboration et al. 2019f), which is also in agreement with close-in-time 3 mm VLBI observations of the M87* jet (Lu et al. 2023).

3.5. Proton-electron temperature coupling

The single-fluid GRMHD simulations trace only the temperature T_i of the dynamically important heavier ions. It is necessary to determine the temperature T_e of the lighter electrons in order to compute the synchrotron emission for the GRRT images. Adopted from Mościbrodzka et al. (2016), the standard models use

$$T_e = \frac{1 + \beta_p^2}{1 + \beta_p^2 R_{high}} T_i . \tag{2}$$

Here, β_p is the ratio of gas over magnetic pressure. In the low β_p outflow regions, an isothermal jet with $T_e \simeq T_i$ forms (compared to other works, we have fixed $R_{low} = 1$ here). The free R_{high} parameter sets $T_e \simeq T_i/R_{high}$ in the accretion disk, where T_i is high due to advective heating and falls off inversely with distance to the black hole. We set $R_{high} = 1, 10, 20, 40, 80, 160$ for M87* and $R_{high} = 1, 10, 40, 160$ for Sgr A*.

3.6. Angular scale

The scale-free GRMHD simulations were evolved on a numerical grid with a characteristic length given by r_g. For the ray-tracing, the mass M and distance D to the black hole were set to $M_{SGRA} = 4.14 \times 10^6 \, M_\odot$, $D_{SGRA} = 8.127$ kpc (Gravity Collaboration et al. 2019; Do et al. 2019; Reid et al. 2019) and $M_{M87} = 6.2 \times 10^9 \, M_\odot$, $D_{M87} = 16.9$ Mpc (Blakeslee et al. 2009; Bird et al. 2010; Gebhardt et al. 2011; Cantiello et al. 2018; Event Horizon Telescope Collaboration et al. 2019f; Broderick et al. 2022; Liepold et al. 2023; Simon et al. 2024) for Sgr A* and M87*, respectively. With these parameters, the characteristic angular scales

$$\vartheta = \frac{GM}{c^2 D} , \tag{3}$$

are $5.0 \, \mu as$ and $3.6 \, \mu as$ for Sgr A* and M87*, respectively.

Following Yao-Yu Lin et al. (2021), we add $\pm 10\%$ random variations to ϑ when generating the synthetic data. These variations reflect uncertainties in M/D and add noise to the data, when they are used for the training of a neural network (Janssen et al. 2025a). Additional noise occurs in the GRRT models, as physical parameters that would otherwise depend on M/D for a fixed flux density (such as optical depth) are not affected when ϑ is varied a posetriori (see the discussions in the appendices of Roelofs et al. 2021; Wong et al. 2022, for example).

3.7. Variability

The GRMHD fluid models were computed with the $t_g = GMc^{-3}$ characteristic time cadence, which is about 20 s for a $4.3 \times 10^6 \, M_\odot$ black hole such as Sgr A* (Gravity Collaboration et al. 2022) and 8 h for a $6.5 \times 10^9 \, M_\odot$ black hole such as M87* (Gebhardt et al. 2011; Event Horizon Telescope Collaboration et al. 2019f). The simulations were evolved until the inner region within about 20 gravitational radii reaches a steady state. In this $t > 10^4 \, t_g$ window, ~ 1000 frames, spaced with a cadence of $5 \, t_g$, were selected to create GRRT images for each model. Hereafter, a single model is referred to as a specific set of GRMHD+GRRT parameters described above for which we have many individual movie image frames.

We do not expect source variability to occur on timescales smaller than $2 \, t_g$ and hence used single GRRT images for full 12 h EHT observing tracks on M87*. For every M87* frame of each model, we generated 10 different realizations of synthetic

Table 1. Antenna parameters used for the synthetic data generation.

Antenna	D (m)	η_{ap}	$\mathcal{S}_{rx}$ (Jy)	$\mathcal{G}_{err}$	$\mathcal{D}$	$\mathcal{P}_{rms}$ (")	$\mathcal{P}_{FWHM}$ (")	$\mathcal{W}$ (mm)	$\mathcal{T}_c$ (s)	P_g (mb)	T_g (K)
ALMA	73	0.73	60	1 ± 0.03	0 ± 0.01	1.0	27	1.5 ± 0.05	10 ± 0.5	555	271
APEX	12	0.63	3300	1 ± 0.05	0 ± 0.01	1.0	27	1.5 ± 0.05	10 ± 0.5	555	271
JCMT	15	0.52	6500	1 ± 0.03	0 ± 0.01	1.0	20	1.5 ± 0.05	5 ± 0.2	626	278
LMT	32	0.31	2400	1 ± 0.08	0 ± 0.03	1.0	10	5.0 ± 0.5	6 ± 0.3	604	275
PV	30	0.43	1000	1 ± 0.03	0 ± 0.03	0.5	11	2.3 ± 0.1	5 ± 0.2	723	270
SMA	16	0.73	3300	1 ± 0.03	0 ± 0.01	1.5	55	1.5 ± 0.05	5 ± 0.2	626	278
SMT	10	0.57	7700	1 ± 0.03	0 ± 0.03	1.0	32	4.0 ± 0.5	5 ± 0.1	695	276
SPT	16	0.73	3300	1 ± 0.03	0 ± 0.03	1.5	55	0.8 ± 0.05	10 ± 0.5	600	270
AMT	15	0.52	7000	1 ± 0.05	0 ± 0.03	1.0	20	10 ± 0.5	5 ± 0.2	772	289

data (see Section 4 below). For Sgr A*, several consecutive image frames were used for the synthetic data generation of a single few minutes-long VLBI scan. To cover a full track, 432 consecutive frames are needed. From the typically 1000 available images, we can therefore select frames 0 to 568 as possible starting points f_0 for a synthetic dataset. From these, we randomly selected 100 different f_0 for each model, each with different realizations of the synthetic data generation parameters. The mass unit of the accretion flow is a free parameter for the GRRT images, determined such that the average flux of a GRMHD run matches the flux measured by the EHT: 0.5 Jansky (Jy) for M87* (Event Horizon Telescope Collaboration et al. 2019c) and 2.4 Jy for Sgr A* (Event Horizon Telescope Collaboration et al. 2022c; Wielgus et al. 2022b).

3.8. Exotic models

3.8.1. M87* Kerr-Newman black hole models

For M87*, we have a set of models based on the Kerr–Newman spacetime metric available. Kerr-Newman spacetimes describe black holes with spin and charge. All models have a SANE magnetic field configuration, with different combinations of positive spin values a_* and dimensionless charges q_* in geometrized units:

– $a_* = 0$ with $q_* = 0$, 0.9375.
– $a_* = 0.25$ with $q_* = 0$, 0.4, 0.9.
– $a_* = 0.4687$ with $q_* = 0$, 0.4, 0.8119.
– $a_* = 0.66$ with $q_* = 0$, 0.33, 0.66.
– $a_* = 0.8$ with $q_* = 0$, 0.2, 0.46875.
– $a_* = 0.9375$ with $q_* = 0$.

Each of the 14 models was then ray-traced in total intensity with $R_{high} = 1, 10, 20, 40, 80, 160$. With 198 frames per model, we have 16,632 images, and for each, multiple M87* synthetic data realizations were created. These models will be described in detail in a future publication (Wondrak et al., in prep.).

3.8.2. Sgr A* dilaton black hole models

Typically, non-GR studies pertinent to the EHT are based on semi-analytical accretion flow models (e.g., Event Horizon Telescope Collaboration et al. 2022e). So far, only a small number of full GRMHD simulations of non-Kerr spacetimes exist (Mizuno et al. 2018; Olivares et al. 2020; Röder et al. 2022, 2023). In this work, we made use of nonrotating dilaton black hole models of Sgr A* as a representative non-GR spacetime.

The dilaton black hole is described by Einstein-Maxwell-Dilaton-Axion (EMDA) gravity, a class of solutions of low-energy

effective string theory (García et al. 1995). The resulting space-time metric is similar to a Schwarzschild metric, except that it is deformed by the dilaton parameter b_*. For more details on EMDA gravity in general, we refer to Wei & Liu (2013), Flathmann & Grunau (2015), and Banerjee et al. (2021a,b).

In this work, the dilaton parameter was fixed at $b_* = 0.504$. That way, the dilaton black hole has the same equatorial innermost stable circular orbit as a Kerr black hole with spin $a_* = 0.6$. This value for b_* is consistent with constraints obtained by Event Horizon Telescope Collaboration et al. (2022e) and Kocherlakota et al. (2021). Two GRMHD simulations with different initial magnetic field configurations, to reach different accretion states, were carried out using the BHAC code (Porth et al. 2017; Mizuno et al. 2018; Röder et al. 2023). Next to SANE states with $\phi_{mag} \sim 1$, accretion flows develop, where ϕ_{mag} comes close to $\phi_{mag} = 10$. The true MAD state is not reached with the initial setup of these simulations.

We generated ray-traced images using the BHOSS code (Younsi et al. 2020) with $R_{high} = 10, 40, 160$ and $i_{los} = 20°, 40°, 60°$. All dilaton images were calculated from purely thermal electron distribution functions in total intensity. From each of the 24 resultant models, 400 Stokes $\mathcal{I}$ frames were generated with a cadence of $10\,t_g$. We used groups of 216 consecutive frames, rotated by $\theta_{PA} = 0°, 30°, 60°, 90°, 120°, 150°,$ and $180°$ to generate multiple Sgr A* synthetic data realizations.

4. Synthetic data generation process

We used the SYMBA pipeline (Roelofs et al. 2020) to compute synthetic visibilities based on GRRT images. SYMBA performs the full forward modeling chain to generate synthetic data that matches the real data as closely as possible (see also Janssen et al. 2022, and references therein). Data corruption effects are added based on first principles with the MEQSILHOUETTE software (Blecher et al. 2017; Natarajan et al. 2022) and the corrupted data are calibrated with the RPICARD pipeline (Janssen et al. 2019b, 2018) in the same way as the real, observational EHT data. We summarize the synthetic data generation parameters for every station in the 2017 EHT array in Table 1. Additionally, we list the Africa Millimeter Telescope (AMT, Backes et al. 2016), with conservative atmospheric parameters, while the exact location for the planned AMT is not yet decided. In this study, we used the location of the 2347m high Gamsberg mountain in Namibia, which is at 23°20′31.9″S 16°13′32.8″E. Results of synthetic data generated with the 2017 EHT array plus AMT are shown in Section 6.

The following subsections provide an overview of all effects that are taken into account for the generation of the synthetic dataset, focusing on the overall parameter space of data corrup-

Table 2. Synthetic data parameter space.

Category	Parameter	Range of values		Parameter space	
		M87*	Sgr A*	M87*	Sgr A*
Standard	ϕ_{mag}	**SANE–MAD**	**SANE–MAD**	**2**	**2**
GRMHD	a_*	**±0.94, ±0.5, 0**	**±0.94, ±0.5, 0**	**10**	**10**
-GRRT	R_{high}	**1, 10, 20, 40, 80, 160**	**1, 10, 40, 160**	**60**	**40**
models	i_{los}	17° (163° if $a_* \geq 0$)	**10, 30, 50, ..., 170°**	60	**360**
	θ_{PA}	288°	**0, 30, 60, ..., 180°**	60	**2,520**
	f_0	0 – 1000	0 – 568	*600,000*	*252,000*
Interstellar	D_{sc}	-	2.7 ± 0.3 kpc	-	*252,000*
scattering	R_{sc}	-	5.4 ± 0.3 kpc	-	*252,000*
screen	$\theta_{\mathrm{sc,maj}}$	-	1.380 ± 0.013 mas	-	*252,000*
toward	$\theta_{\mathrm{sc,min}}$	-	0.703 ± 0.013 mas	-	*252,000*
Sgr A*	$\phi_{\mathrm{sc,PA}}$	-	81.9° ± 0.2°	-	*252,000*
	α_{sc}	-	1.38 ± 0.06	-	*252,000*
	$r_{\mathrm{sc,in}}$	-	800 ± 200 km	-	*252,000*
Data	$\mathcal{P}_{\mathrm{rms}}$	0.5″ – 1.5″		*600,000*	*252,000*
corruption	$\mathcal{W}$	(0.8 – 4.0) ± (0.05 – 0.5) mm		*600,000*	*252,000*
effects	$\mathcal{T}_{\mathrm{c}}$	(5 – 10) ± (0.1 – 0.5) s		*600,000*	*252,000*
	D_ϕ	$\propto \mathcal{T}_{\mathrm{c}}^{-5/3}$		*600,000*	*252,000*
	$\mathcal{D}$	1 – 3 %		*600,000*	*252,000*
	$\mathcal{G}_{\mathrm{err}}$	3 – 8 %		*600,000*	*252,000*
	σ_{th}^2	$\propto \mathrm{SEFD}_1 \times \mathrm{SEFD}_2$		*600,000*	*252,000*
		Total number of datasets:		600,000	252,000

Notes. – Parameter space values that are drawn together (f_0, interstellar scattering, and data corruption effects) are written in italics. Non-nuisance parameters of physical interest are marked in bold. The scattering parameters are taken from Table 3 in Johnson et al. (2018). The GRMHD-GRRT parameter space is different for the exotic models (see Section 3.8).

tion and calibration processes. More details about the emulated effects, their impact on the data, and justifications for the range of parameters chosen are given in Roelofs et al. (2020). The M87* and Sgr A* data are processed in the same way, unless stated otherwise.

4.1. Interstellar scattering

Turbulent magnetized plasma along the line of sight toward the galactic center causes a scattering of radio waves emitted by Sgr A* that are received at Earth (e.g., Davies et al. 1976; van Langevelde et al. 1992; Bower et al. 2014; Dexter et al. 2017). The scattering screen results in an angular broadening of the image structure. Additionally, density irregularities in the plasma, that may move across the line of sight, produce variable substructures in the image.

For Sgr A* synthetic data, we added interstellar scattering data corruption effects from the Psaltis et al. (2018) and Johnson et al. (2018) model. A random realization with values drawn from the parameters listed in Table 2 was created for each observation. More details about the scattering screen toward Sgr A* can be found in Section 5.1.2 of Event Horizon Telescope Collaboration et al. (2022b).

4.2. Baseline coverage

We generated the M87* and Sgr A* synthetic data with the (u, v) coverage (Figure 2) of the processed 2017 EHT observations of M87* on 11 April and Sgr A* on 7 April (Event Horizon Telescope Collaboration et al. 2019c, 2022b), respectively. Fringe

non-detections and other data dropouts were thereby taken into account.

For the Sgr A* observations, the ALMA, Atacama Pathfinder Experiment (APEX), James Clerk Maxwell Telescope (JCMT), Large Millimeter Telescope Alfonso Serrano (LMT), IRAM 30 m Telescope (PV), Submillimeter Array (SMA), Submillimeter Telescope (SMT), and South Pole Telescope (SPT) observatories participated in the observations. M87* was observed by the same telescopes, minus the SPT, which cannot see the source.

4.3. Antenna pointing errors

We used an ad hoc prescription to model antenna pointing errors. Telescope beam offsets ρ are drawn from a normal distribution centered around zero with a standard deviation of $\mathcal{P}_{\mathrm{rms}}$. These offsets introduce stochastic variability on a specified atmospheric coherence time $\mathcal{T}_{\mathrm{c}}$ and gross per-scan amplitude losses, for which ρ slowly increases by 3 % per scan until new pointing offsets are drawn every $\sim$ 5 scans (indicating that telescopes have obtained new pointing solutions for their dishes during the observing run).

We assumed Gaussian telescope beams with a full-width at half maximum of $\mathcal{P}_{\mathrm{FWHM}}$, for which fractional amplitude losses Δz are given by

$$\Delta z = \exp\left(-4\ln 2 \frac{\rho^2}{\mathcal{P}_{\mathrm{FWHM}}^2}\right). \tag{4}$$

4.4. Earth atmosphere

We use the atmospheric module of MeqSilhouette to simulate four data corruption effects:
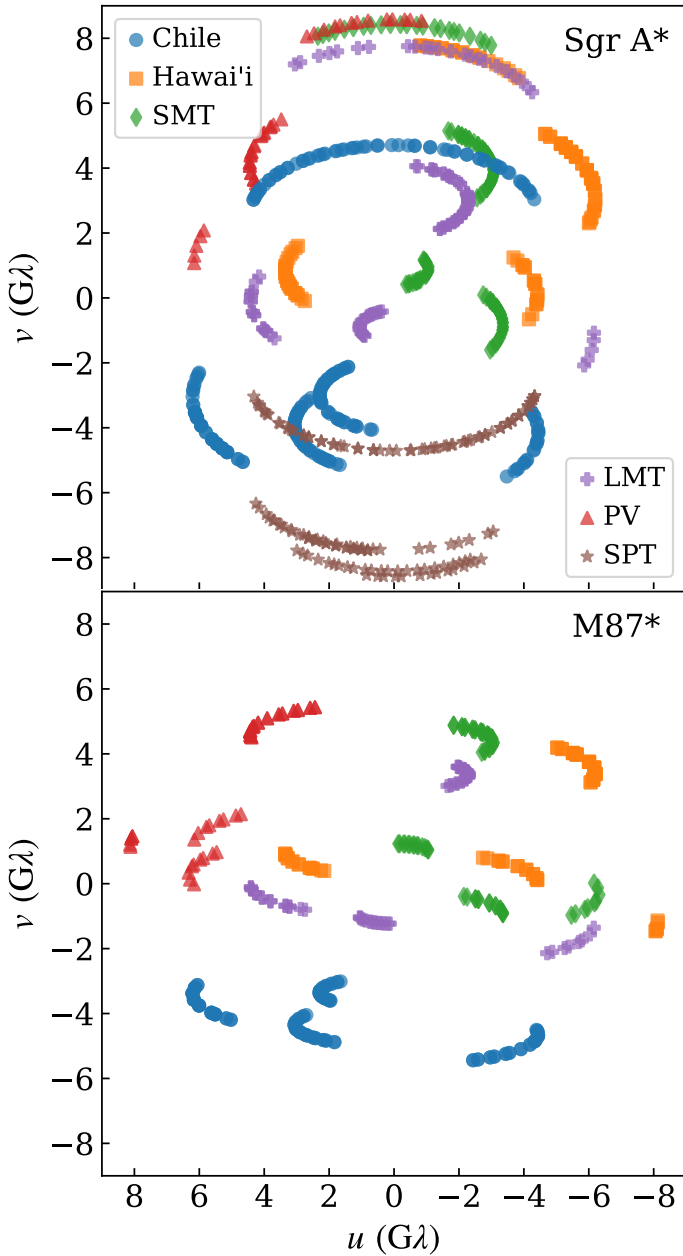
Fig. 2. Baseline coverage of the 7 April Sgr A* (top) and 11 April 2017 M87* (bottom) 226.1 – 228.1 GHz EHT data processed with Rpicard. The Chile and Hawai'i markers encompass baselines to the co-located ALMA–APEX and JCMT–SMA stations, respectively. The data are averaged over VLBI scan durations and over all frequency channels here and the zero-spacings between co-located sites are not plotted. Conjugate baseline pairs (1-2 and 2-1) are displayed differently following the legends shown in the upper panel.

1. The Atmospheric Transmission at Microwaves (ATM) software (Pardo et al. 2001) is used to solve the radiative transfer equation and to compute the amount of signal attenuation along the line of sight through the Earth's atmosphere. ATM is initialized based on the amount of precipitable water vapour $\mathcal{W}$ at zenith, the local pressure P_g, and temperature T_g at each antenna. The values for these quantities were taken from EHT observations logged in the VLBIMONITOR (Event Horizon Telescope Collaboration et al. 2019b).

2. ATM also computes the amount of sky noise T_{sky} arising from the radiation produced by the atmosphere at the observing frequency.

3. Wet dispersive path delays are simulated with ATM.

4. As described in Section 3.2.2 of Natarajan et al. (2022), Kolmogorov phase turbulence $\delta\phi$ from the troposphere is simulated with a power law approximation of the phase structure function $D_\phi \propto \mathcal{T}_c^{-5/3}$ for a zenith atmospheric coherence time $\mathcal{T}_c$. The time series of phase errors are $\propto \mu\nu\,\delta\phi$, given the airmass μ toward the horizon and observing frequency ν. We simulated visibilities with 32 channels, spanning a bandwidth of 2 GHz around a central frequency of 227 GHz, where the ray-tracing was performed to make the GRRT images. Compared to previous simulation data from MeqSilhouette, where $\delta\phi$ itself was also approximated with a power-law (Blecher et al. 2017), we now use a Cholesky factorization in MeqSilhouette v2 (Natarajan et al. 2022).

4.5. Thermal noise

Thermal noise is determined by the System Equivalent Flux Densities (SEFDs) of every antenna, which measure the total noise contribution along the signal path. We computed the SEFDs as

$$\mathrm{SEFD} = \mathcal{S}_{rx} + 8\frac{k_B T_{sky}}{\pi\eta_{ap}D^2}, \tag{5}$$

with $\mathcal{S}_{rx}$ the noise contribution from the telescope's receiver, k_B the Boltzmann constant, D the diameter of the telescope dish, and η_{ap} the aperture efficiency.

Noise was added to the visibilities by randomly drawing from a Gaussian probability distribution function with a standard deviation of

$$\sigma_{th} = \frac{1}{\eta_Q}\sqrt{\frac{\mathrm{SEFD}_1\mathrm{SEFD}_2}{2\,\mathrm{GHz}\times 0.5\,\mathrm{s}}}. \tag{6}$$

Here, $\eta_Q = 0.85457$ is the EHT quantization efficiency for the CASA-based EHT data reduction.

4.6. Polarization leakage

We modeled residual polarization leakage ($\mathcal{D}$-term) effects based on the accuracy at which we can constrain their magnitude for the EHT (Event Horizon Telescope Collaboration et al. 2021). We assume the $\mathcal{D}$-terms to be constant over observing tracks as well as the frequency bandwidth and add noise by randomly drawing from a Gaussian probability distribution function with standard deviation $\sigma_{\mathcal{D}}$ for each station. For the co-located ALMA, APEX and JCMT, SMA sites, we used $\sigma_{\mathcal{D}} = 0.01$. For all other antennas, where the leakages are more difficult to constrain (see, e.g., Issaoun et al. 2022; Jorstad et al. 2023), we used $\sigma_{\mathcal{D}} = 0.03$.

4.7. Telescope amplitude gain errors

In addition to antenna pointing errors, we simulated gross gain errors $\mathcal{G}_{err}$ for each station. These arise mostly from gain measurement errors in the real data. Typical gain errors are $\mathcal{O}(3\,\%)$ at a 1-sigma level (Janssen et al. 2019a). The LMT has the largest gain uncertainties of $\sim 8\,\%$. The gains are mostly dependent between the RCP and LCP circular polarization feed signal paths of the telescopes. We based LCP gain offsets on a random draws of RCP offsets with an additional $\mathcal{O}(1\,\%)$ random relative offset (Section 2.2). Examples of residual self-calibration gains from the imaging of EHT data are given in Table 14 of Event Horizon Telescope Collaboration et al. (2019d).

4.8. Phase calibration

The path delay and phase turbulence corruption effects added by the simulated Earth atmosphere (Section 4.4) have to be corrected to obtain a realistic dataset. The total flux density and source structure of the input model image determine the correlated flux density across the interferometric baselines and thereby how well these effects can be corrected in the forward modeling. Fringe-fitting (Thompson et al. 2017) the data to solve for delay offsets determines if the source is detected for particular scans and antennas. Intra-scan phase turbulence is corrected by fringe-fitting segmented pieces of data. The length of the segmentation intervals is determined by the S/N on baselines to a chosen reference station (Janssen et al. 2019b). Akin to the real data, this left residual atmospheric phase wander in low S/N data (on long baselines or near interferometric nulls), which causes decoherence effects when the data are averaged in time.

4.9. Amplitude calibration

Accurate a priori models for the flux densities of resolved and variable VLBI sources are typically not available. Commonly, the amplitude calibration is done with a priori estimations for each station's total noise budget instead (Equation 5). For continuum high-frequency VLBI observations, the sky temperature is the only substantial variable source of noise. In good weather conditions, T_{sky} only varies with the amount of arimass toward the horizon, which is set by the telescope elevation angle. Usually, single load chopper calibration measurements are performed by every station right before any VLBI scan to measure the total noise and atmospheric attenuation. We emulated these measurements by correcting the ATM attenuation τ of every scan with the value of τ given at the start of the scan, leaving intra-scan variations of the attenuation due to small changes in elevation angle uncorrected. Higher-order noise contributions, which could arise due to spillover at the telescope or the contribution from the astronomical source itself, for example, are ignored here.

Additionally, the network-calibration technique (Blackburn et al. 2019) is used in the real data to accurately constrain the gains of co-located sites based on ALMA and SMA measurements of the total source flux density at large scales. The same flux is assumed to be measured on the short ALMA-APEX and JCMT-SMA EHT baselines, both of which do not resolve structures smaller than ∼100 mas. For the time-variable Sgr A* models, a time-dependent light curve (Wielgus et al. 2022b; Event Horizon Telescope Collaboration et al. 2022b) is constructed and used to constrain gains at the time cadence of the individual input model movie frames, mimicking the network-calibration employed for the observational Sgr A* data. For the synthetic data, we assumed that ALMA calibration errors are negligible for the network calibration, that is, we used the uncorrupted model fluxes.

5. Data workflow, content, format, and availability

The input models and output synthetic datasets are staged on the CyVerse (formerly the iPlant Collaborative) data storage (Goff et al. 2011; Merchant et al. 2016) and the synthetic data generation software SYMBA is run on the Open Science Grid (Pordes et al. 2007; Sfiligoi et al. 2009; OSG 2006, 2015) as well as the Cobra HPC system at the Max Planck Computing and Data Facility. The PEGASUS workflow management system (Deelman et al. 2015) is used to schedule the computations and handle the data exchange between CyVerse and the Open Science Grid in a reproducible

manner. Access to the initial model data and final synthetic data on CyVerse is facilitated through `iRODS`.[5]

We simulated data for all four correlation products. For circular feeds, we have the RR, LL parallel-hand and RL, LR cross-hand complex visibilities measured on all baselines (except for the JCMT, which only observed single-pol data in 2017). The Stokes parameters can be extracted from these measurements as

$$\mathcal{I} = \frac{1}{2}(\text{RR} + \text{LL}) \tag{7}$$

$$\mathcal{Q} = \frac{1}{2}(\text{RL} + \text{LR}) \tag{8}$$

$$\mathcal{U} = \frac{i}{2}(\text{LR} - \text{RL}) \tag{9}$$

$$\mathcal{V} = \frac{1}{2}(\text{RR} - \text{LL}) \,, \tag{10}$$

with $i = \sqrt{-1}$.

SYMBA creates frequency-resolved synthetic data in the `MeasurementSetv2`[6] format. We do not store these datasets permanently; instead, we store the frequency-averaged data in the `UVFITS`[7] format on CyVerse. We have used the `49a813d2dc62eac809f3909bee0d38a8b113ffc4`[8] SYMBA DOCKER[9] container to generate the synthetic date presented in this work.

As our final data product, we bundled the complex correlation coefficients for groups of 1000 synthetic datasets together, each labeled with ϕ_{mag}, a_*, R_{high}, i_{los}, and θ_{PA} in single `TFRecord`[10] files. These files efficiently serialize structured data in a coding language- and hardware platform-independent manner based on Google's `protocol buffer`[11] method. The self-contained `TFRecords` do not allow for random data access but take up less disk space, are easier to handle compared to the large number of individual `UVFITS` files, allow for rapid parallel I/O operations, and consistently store data together with their labels. Machine learning frameworks such as TENSORFLOW (Abadi et al. 2016a,b) can combine any number of individual `TFRecord` files into single objects, for which data can be loaded in optimally sized chunks based on the amount of available system memory and operations such as shuffling and batching are easily performed. The real and imaginary components of the visibilities are stored in one-dimensional arrays sorted by (u, v)-distance for each correlation product. In cases where fringe non-detections occur in the forward modeling process, the flagged visibilities are replaced with zeroes.

The single data set `UVFITS` files have a size of 652 Kilobytes and 1.5 Megabytes for M87* and Sgr A*, respectively. The corresponding `TFRecord` file sizes of 1000 bundled datasets are 168 and 423 Megabytes, respectively. For the full parameter space shown in Table 2, there are about 100 Gigabytes of `TFRecord` data for each source. Access to the `UVFITS` and `TFRecord` files will be given upon reasonable request. The availability of the
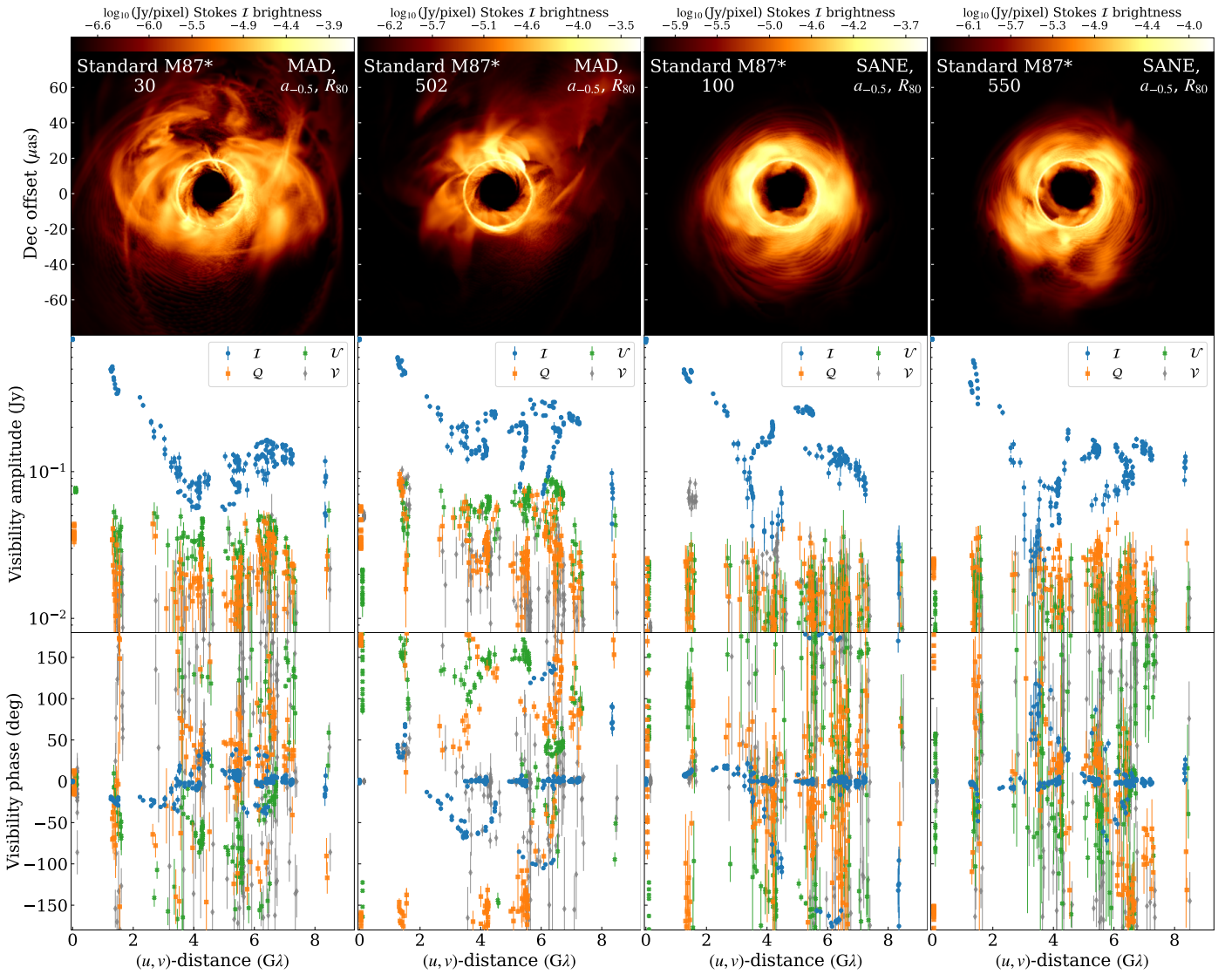

Fig. 3. Four synthetic datasets based on two realizations of two standard M87* models are presented. GRRT frame numbers are displayed in the top left corners. The top row shows the total intensity ray-traced ground-truth model images on logarithmic scales with varying dynamic ranges. Visibility amplitudes on a logarithmic scale and phases of corresponding synthetic data realizations are displayed with thermal noise error bars as a function of baseline length in the middle and bottom rows, respectively. The measurements shown can come from different orientations at the same baseline length. For better readability, the visibilities have been averaged over scan durations, amplitudes lower than 0.008 Jy have been clipped, and the values of the different Stokes parameters are each offset by 50 Mλ on the x-axis. Spin $a_* = s$ and $R_{\text{high}} = r$ parameters are listed in a shorthand notation as a_s and R_r in the top-right corner of each model image.

observational EHT data and corresponding processing software is described in Janssen et al. (2025b).

6. Synthetic data features

Figure 3 depicts example synthetic data realizations for four different M87* GRRT frames with $a_* = -0.5$ and $R_{\text{high}} = 80$. The two realizations of the more variable MAD model show more extended emission compared to the two SANE realizations. The synthetic visibilities displayed here are averaged over VLBI scan durations and baseline orientations. We compared the differences between the data realizations of the same model with synthetic data, where thermal noise was added as the only data corruption effect. We found that the differences in the total intensity data are primarily caused by intrinsic model variations, while differences in polarization-sensitive visibilities are mostly due to differences in the simulated data corruption effects, the residual $\mathcal{D}$-terms

for example. The strong circular polarization amplitude at about 1.8 Gλ in the SANE GRRT # 100 model realization for instance is the result of telescope gain errors. For the Figure 3 case study, the presence of a prominent total intensity visibility minimum at about 3.5 Gλ is a distinctive data feature only present in the SANE model. We show and discuss the differences between the Symba simulated synthetic data and ground truth from the underlying model image in Appendix A.

Figure 4 presents example synthetic data realizations from the highly variable Sgr A* models; here two MAD and two SANE realizations with $a_* = 0$, $R_{\text{high}} = 1$, $i_{\text{los}} = 90°$, and $\theta_{\text{PA}} = 0$. These models show edge-on disk-dominated synchrotron emission in a Schwarzschild spacetime. In this case, salient features can be found in the displayed linear polarization visibility phases. Compared to the SANE model, the MAD's Q and U phases have a distinct structure and are more coherent with baseline length.
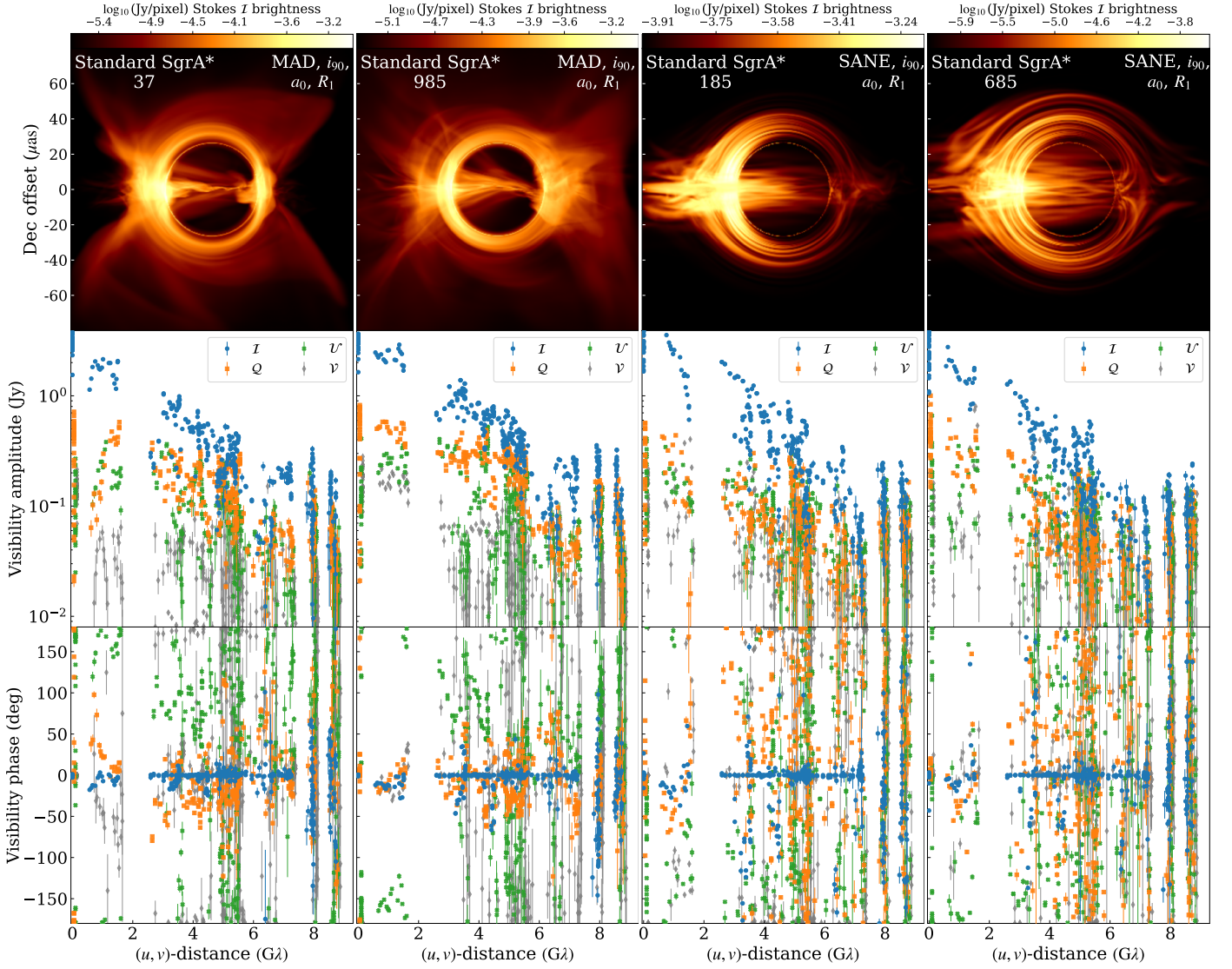
Fig. 4. Same as Figure 3 but for $\theta_{PA} = 0$ standard Sgr A* models, where the $i_{los} = l$ parameter is indicated with a i_l label. Single representative GRRT frames are shown for the data that are built up from a movie of many consecutive frames. Zero-baseline fluxes above 3.9 Jy have been clipped.

Figure 5 shows synthetic data for alternative Sgr A* and M87* models (Section 3.8) for both the 2017 baseline coverage as well as possible future EHT observations, where the AMT or ngEHT array (Doeleman et al. 2019) join the EHT. For Sgr A* EHT observations, the AMT adds substantial improvements to the baseline coverage (La Bella et al. 2023). Currently, the longest baseline on our primary targets is PV–SPT with 8.7 Gλ at 230 GHz. With the AMT, we will have SMT–AMT out to 8.9 Gλ at 230 GHz with an improved northeast and southwest resolution. To demonstrate the capabilities of a significantly enhanced EHT array, comparable to the 23 stations that can currently observe at 86 GHz with the Global mm-VLBI Array (GMVA, Krichbaum et al. 2006), we have simulated observations of 22 dishes from the combined current EHT plus full ngEHT reference array 1 as described in Roelofs et al. (2023) and Doeleman et al. (2023). The 22 mm VLBI sites will yield a very dense (u, v) coverage and image reconstructions with much better dynamic ranges. Since the GRRT images of our exotic models were ray-traced only in total intensity, we focus only on the Stokes $\mathcal{I}$ data here.

For the dilaton models, a noticeable feature is the structural and total flux variability excesses of the SANE model over the MAD model. This feature can be used to easily distinguish the

two $b_* = 0.504$, $R_{high} = 40$, $i_{los} = 20°$, $\theta_{PA} = 0$ models through the Stokes $\mathcal{I}$ measurements. Typically MAD simulations are accompanied by more violent outflow eruptions and are thus overall more variable than SANE models, at least for the standard models considered in this work. As the nonrotating dilaton models stay below $\phi_{mag} < 10$, no strong eruptions occur. The variability excess of the SANE dilaton model will be investigated in a future study.

The two SANE $a_* = 0.25$, $R_{high} = 80$, M87* Kerr-Newman models in Figure 5 differ in their charge q_* parameters – one model has $q_* = 0$ and the other has $q_* = 0.9$. Both models are similarly compact as the standard SANE models shown in Figure 3, which also have $R_{high} = 80$. Given that the size of the event horizon is $1 + \sqrt{1 - a_*^2 - q_*^2}$, the size of the photon ring is noticeably smaller for the $q_* = 0.9$ model. The $q_* = 0$ model shows some foreground emission with a low surface brightness inside of the photon ring. For the $q_* = 0.9$ model, there is a bright ring of foreground emission inside of the photon ring. As a result, the black hole shadow appears to be substantially smaller. The smaller ring size caused by the black hole charge is clearly identifiable in the shift of the total intensity visibility minima location toward longer baseline distances.
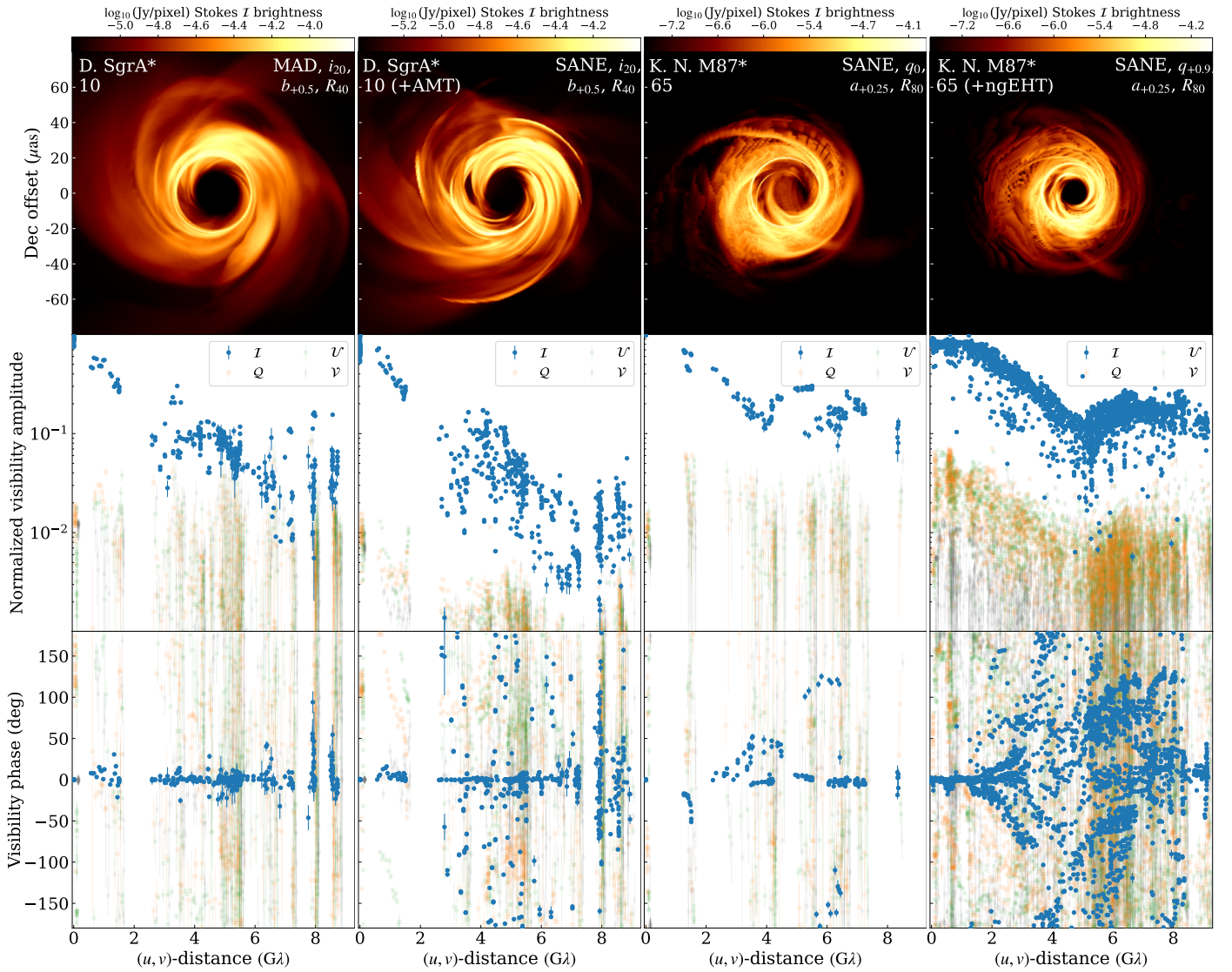
Fig. 5. Similar to Figure 3 and Figure 4 for the M87* and Sgr A* models, respectively. Here, we have Sgr A* dilaton (D.) and M87* Kerr-Newman (K. N.) models (Section 3.8). For the former, the dilaton parameter $b_* = 0.504$ (see text) is labeled as $b_{+0.5}$. For the latter, the geometrized charge $q_* = C$ is given with a q_C label. As the D. and K. N. models were ray-traced only in total intensity, we have faded out the Stokes Q, U, V data. Visibility amplitudes have been normalized to unity. Next to the 2017 coverage, measurements are shown for possible future EHT observations, where the AMT or ngEHT would join (see text).

Building upon earlier horizon-scale GRMHD-GRRT variability studies of M87* by Satapathy et al. (2022) and Sgr A* by Georgiev et al. (2022), we analyze the influence of model variability on parameter inference, taking the full forward modeling chain for the synthetic data generation into account for the first time. We note here that telescope-based gain errors and inaccuracies in the signal stabilization method impact closure quantities when data are averaged in time or frequency. Figure 6 shows closure phase variability from three standard M87* models on the ALMA-LMT-SMT and LMT-PV-SMT triangles, which show little variability in the observational data over the six-day-long extend of the 2017 EHT observations (Satapathy et al. 2022). Different realizations that are based on the same model image have comparatively little influence on the M87* closure phase variability for the two triangles analyzed here. The dominant source of variability comes from the differences between individual model frames, which is most evident for the ALMA-LMT-SMT data of the SANE, $a_* = -0.94$, $R_{high} = 160$ model.

The LMT-PV-SMT data do not have discriminative power for the three M87* models shown. With only a few observations, the ALMA-LMT-SMT data alone cannot be used to distinguish between the models either. With long-term monitoring, the model variability can be used to distinguish between the example parameters here; the one SANE model differs substantially from the two MADs based on the overall degree of variability. The MAD models themselves differ only by the black hole spin. Their degree of variability is comparable, but the median closure phases value, measured over years, is lower for the $a_* = 0$ model compared to the $a_* = 0.94$ one.

These simple examples make use of only a small part of the simulation model parameter space and features in the data. In Janssen et al. (2025a), we study how well a Bayesian neural network can distinguish standard models when the full parameter space and all data products are taken into account. We find that our neural network, by using the full information content of the data, is able to distinguish all model parameters in the presence of the specific closure phase degeneracies presented here. Furthermore,
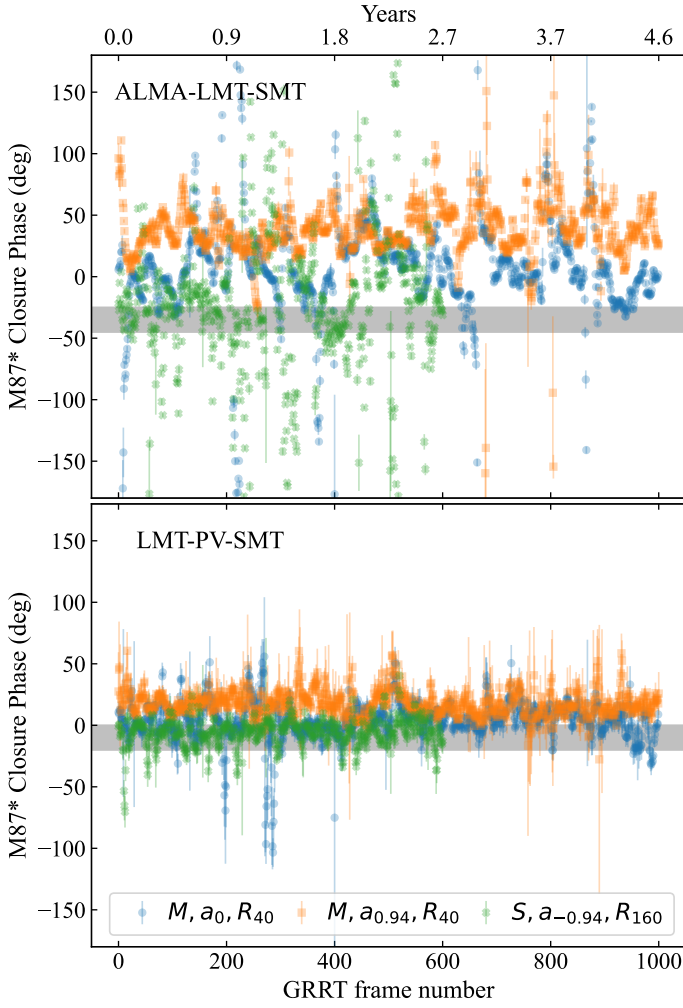
Fig. 6. Total intensity closure phase evolution of M87* synthetic data from example MAD (*M*) and SANE (*S*) standard models as a function of model variability for the ALMA-LMT-SMT and LMT-PV-SMT triangles, respectively. Spin $a_* = s$ and $R_{high} = r$ parameters are listed in a shorthand notation as a_s and R_r in the legend of the bottom panel. Each data point corresponds to a VLBI scan-averaged synthetic data closure phase measured at 02:27 UT on 11 April 2017 and the standard deviations plotted are computed from multiple synthetic data realizations of the same model frame. The corresponding measurement range from the 2017 observational data is depicted with a gray band. The SANE model shown here has been ray-traced for 600 frames only.
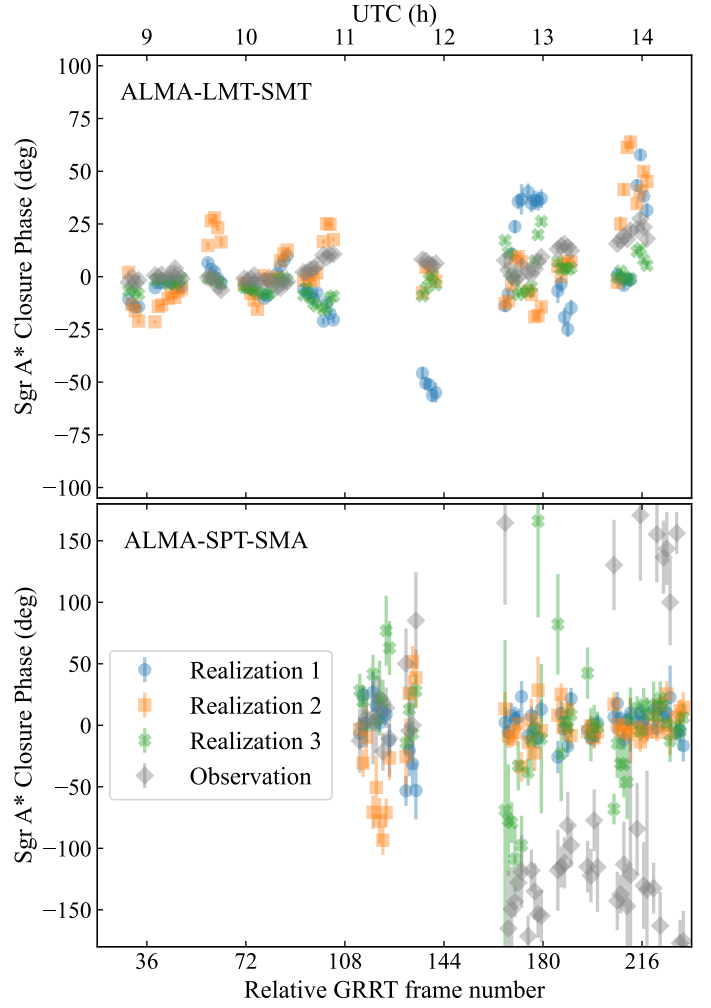
Fig. 7. ALMA-LMT-SMT and ALMA-SPT-SMA total intensity closure phase evolution of 120 s averaged Sgr A* data over the course of the 2017 EHT observing track on 7 April. Three synthetic data realizations of the variable MAD, $a_* = 0.5$, $R_{high} = 160$, $i_{los} = 30°$, $\theta_{PA} = 0$ model are shown next to the observational data. The error bars depict the a priori thermal noise estimations of the data.

variability is not taken into account. In Janssen et al. (2025a), we show that a suitably trained Bayesian deep neural network can accurately fit parameters of interest, even when different model realizations exhibit significant variations in specific data products.

7. Summary and conclusions

In preparation for machine-learning-based GRMHD-GRRT parameter inference with EHT data (Janssen et al. 2025a), we produced a comprehensive synthetic data library matching different millimeter VLBI observations. Most of our 962,000 datasets are made to match the 2017 EHT observations of Sgr A* and M87* on 7 and 11 April, respectively. Additionally, we studied synthetic data from possible future observations of the EHT+AMT and EHT+ngEHT. Next to a standard set of Kerr black hole models, that broadly sample the ϕ_{mag}, a_*, R_{high}, and i_{los} parameter space, we also included Kerr-Newman and EMDA gravity black hole solutions in our library.

We presented a substantial upgrade in the CASA-based EHT data reduction pathway, obtaining the hitherto highest quality EHT data. The calibration process with its upgrades presented here is taken into account for our generation of realistic synthetic

the network is robust against the intrinsic model variations – a network trained only on the first half of GRRT frames can accurately predict the parameters of the latter half of frames for each model.

Figure 7 shows ALMA-LMT-SMT and ALMA-SPT-SMA closure phase variations for three synthetic data realizations of the MAD $a_* = 0.5$, $R_{high} = 160$, $i_{los} = 30°$, $\theta_{PA} = 0$, standard Sgr A* model for the 7 April 2017 EHT track. Due to the short gravitational timescale of Sgr A*, multiple GRRT frames make up a single synthetic dataset, resulting in significant differences concerning how well different realizations of the same model agree with the observational data. This effect is most evident in the ALMA-LMT-SMT data. For ALMA-SPT-SMA, we found that only a few model frames agree with the observational measurements after 12 UT. Consequently, the MAD $a_* = 0.5$, $R_{high} = 160$, $i_{los} = 30°$ model was found to pass the EHT data constraints used in Event Horizon Telescope Collaboration et al. (2022f), when

data. Thereby, direct comparisons of the raw visibilities can be made between the observational and synthetic data.

We analyzed the synthetic data from several selected models for dedicated case studies. We also demonstrated that a long-term monitoring of M87* is needed to discriminate between different models through closure phase measurements. For the $a_* = -0.5$ and $R_{high} = 80$ M87* model, we found that intrinsic model variability has a strong imprint on the total intensity visibility data, while variations in polarized light are mostly caused by data corruption effects. For the $a_* = 0$, $R_{high} = 1$, $i_{los} = 90°$, $\theta_{PA} = 0$, Sgr A* model on the other hand, the linear polarization phases are well suited for distinguishing between ϕ_{mag} accretion states. We showed that the MAD $a_* = 0.5$, $R_{high} = 160$, $i_{los} = 30°$, $\theta_{PA} = 0$, Sgr A* model, which was found to pass the stationary EHT data constraints considered in Event Horizon Telescope Collaboration et al. (2022f), does not actually fit the measured ALMA-SPT-SMA closure phase. It is clear that different models have different salient features that can be used as discriminating factors.

For our Sgr A* dilaton models, we found the intrinsic model variability in total intensity within a single EHT observing track to be a good indicator for ϕ_{mag}. Surprisingly, the variability is higher for SANE models, opposite to the typical behavior of the standard GRMHD-GRRT models. For the SANE $a_* = 0.25$, $R_{high} = 80$, Kerr-Newman M87* model, the possible presence of a large enough black hole charge is easily identifiable by the shrinking black hole shadow size.

The case studies presented in this work are useful to obtain intuition for the EHT observational and model data. Equipped with the gained insights, we have performed parameter inference studies with machine learning methods in Janssen et al. (2025a) and Janssen et al. (2025b).

References

Abadi, M., Agarwal, A., Barham, P., et al. 2016a, arXiv e-prints, arXiv:1603.04467

Abadi, M., Barham, P., Chen, J., et al. 2016b, arXiv e-prints, arXiv:1605.08695

Backes, M., Müller, C., Conway, J. E., et al. 2016, in The 4th Annual Conference on High Energy Astrophysics in Southern Africa (HEASA 2016), 29

Baczko, A.-K., Kadler, M., Ros, E., et al. 2024, A&A, 692, A205

Balbus, S. A. & Hawley, J. F. 1998, Reviews of Modern Physics, 70, 1

Banerjee, I., Mandal, B., & SenGupta, S. 2021a, MNRAS, 500, 481

Banerjee, I., Mandal, B., & SenGupta, S. 2021b, Phys. Rev. D, 103, 044046

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

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

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

Blandford, R. D. & Znajek, R. L. 1977, MNRAS, 179, 433

Blecher, T., Deane, R., Bernardi, G., & Smirnov, O. 2017, MNRAS, 464, 143

Bower, G. C., Deller, A., Demorest, P., et al. 2014, ApJ, 780, L2

Broderick, A. E., Pesce, D. W., Gold, R., et al. 2022, ApJ, 935, 61

Cantiello, M., Blakeslee, J. P., Ferrarese, L., et al. 2018, ApJ, 856, 126

CASA Team, Bean, B., Bhatnagar, S., et al. 2022, PASP, 134, 114501

Cui, Y., Hada, K., Kawashima, T., et al. 2023, Nature, 621, 711

Davies, R. D., Walsh, D., & Booth, R. S. 1976, MNRAS, 177, 319

Deelman, E., Vahi, K., Juve, G., et al. 2015, Future Generation Computer Systems, 46, 17, funding Acknowledgements: NSF ACI SDCI 0722019, NSF ACI SI2-SSI 1148515 and NSF OCI-1053575

Dexter, J., Deller, A., Bower, G. C., et al. 2017, MNRAS, 471, 3563

Dexter, J., Scepi, N., & Begelman, M. C. 2021, ApJ, 919, L20

Do, T., Hees, A., Ghez, A., et al. 2019, Science, 365, 664

Doeleman, S., Blackburn, L., Dexter, J., et al. 2019, in Bulletin of the American Astronomical Society, Vol. 51, 256

Doeleman, S. S., Barrett, J., Blackburn, L., et al. 2023, Galaxies, 11, 107

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022a, ApJ, 930, L12

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022b, ApJ, 930, L13

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022c, ApJ, 930, L14

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022d, ApJ, 930, L15

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022e, ApJ, 930, L17

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJ, 875, L1

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019b, ApJ, 875, L2

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019c, ApJ, 875, L3

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019d, ApJ, 875, L4

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019e, ApJ, 875, L5

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019f, ApJ, 875, L6

Event Horizon Telescope Collaboration: Akiyama, K., Alberdi, A., et al. 2021, ApJL, 910, 48

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022f, ApJ, 930, L16

Fishbone, L. G. & Moncrief, V. 1976, ApJ, 207, 962

Flathmann, K. & Grunau, S. 2015, Phys. Rev. D, 92, 104027

García, A., Galtsov, D., & Kechkin, O. 1995, Phys. Rev. Lett., 74, 1276

García, A., Galtsov, D., & Kechkin, O. 1995, Phys. Rev. Lett., 74, 1276

Gebhardt, K., Adams, J., Richstone, D., et al. 2011, ApJ, 729, 119

Georgiev, B., Pesce, D. W., Broderick, A. E., et al. 2022, ApJ, 930, L20

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

Goff, S. A., Vaughn, M., McKay, S., et al. 2011, Frontiers in Plant Science, 2, 34

Gold, R., Broderick, A. E., Younsi, Z., et al. 2020, ApJ, 897, 148

Gravity Collaboration, Abuter, R., Aimar, N., et al. 2022, A&A, 657, L12

Gravity Collaboration, Abuter, R., Amorim, A., et al. 2019, A&A, 625, L10

Gross, D. J., Harvey, J. A., Martinec, E., & Rohm, R. 1985, Phys. Rev. Lett., 54, 502

Ho, L. C. 2008, ARA&A, 46, 475

Hoak, D., Barrett, J., Crew, G., & Pfeiffer, V. 2022, Galaxies, 10, 119

Issaoun, S., Wielgus, M., Jorstad, S., et al. 2022, ApJ, 934, 145

Janssen, M., Blackburn, L., Issaoun, S., et al. 2019a, EHT Memo Series, 2019-CE-01 (https://eventhorizontelescope.org/for-astronomers/memos)

Janssen, M., Chan, C.-k., Davelaar, J., et al. 2025a, A&A, appected (Paper II)

Janssen, M., Chan, C.-k., Davelaar, J., et al. 2025b, A&A, appected (Paper III)

Janssen, M., Falcke, H., Kadler, M., et al. 2021, Nature Astronomy, 5, 1017

Janssen, M., Goddi, C., Falcke, H., et al. 2018, in 14th European VLBI Network Symposium & Users Meeting (EVN 2018), 80

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

Janssen, M., Radcliffe, J. F., & Wagner, J. 2022, Universe, 8, 527

Johnson, M. D., Narayan, R., Psaltis, D., et al. 2018, ApJ, 865, 104

Jorstad, S., Wielgus, M., Lico, R., et al. 2023, ApJ, 943, 170

Kerr, R. P. 1963, Phys. Rev. Lett., 11, 237

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

Kocherlakota, P., Rezzolla, L., Falcke, H., et al. 2021, Phys. Rev. D, 103, 104047

Krichbaum, T. P., Agudo, I., Bach, U., Witzel, A., & Zensus, J. A. 2006, in Proceedings of the 8th European VLBI Network Symposium, 2

La Bella, N., Issaoun, S., Roelofs, F., Fromm, C., & Falcke, H. 2023, A&A, 672, A16

Liepold, E. R., Ma, C.-P., & Walsh, J. L. 2023, ApJ, 945, L35

Liska, M., Hesp, C., Tchekhovskoy, A., et al. 2021, MNRAS, 507, 983

Lu, R.-S., Asada, K., Krichbaum, T. P., et al. 2023, Nature, 616, 686

Lynden-Bell, D. 1969, Nature, 223, 690

Merchant, N., Lyons, E., Goff, S., et al. 2016, PLOS Biology, 14, e1002342

Mertens, F., Lobanov, A. P., Walker, R. C., & Hardee, P. E. 2016, A&A, 595, A54

Mizuno, Y. 2022, Universe, 8, 85

Mizuno, Y., Younsi, Z., Fromm, C. M., et al. 2018, Nature Astronomy, 2, 585

Mościbrodzka, M., Falcke, H., & Shiokawa, H. 2016, A&A, 586, A38

Mościbrodzka, M. & Gammie, C. F. 2018, MNRAS, 475, 43

Natarajan, I., Deane, R., Martí-Vidal, I., et al. 2022, MNRAS, 512, 490

Newman, E. T., Couch, E., Chinnapared, K., et al. 1965, Journal of Mathematical Physics, 6, 918

Olivares, H., Younsi, Z., Fromm, C. M., et al. 2020, MNRAS, 497, 521

OSG. 2006, https://doi.org/10.21231/906P-4D78

OSG. 2015, https://osdf.osg-htc.org/

Paraschos, G. F., Kim, J. Y., Wielgus, M., et al. 2024, A&A, 682, L3

Pardo, J. R., Cernicharo, J., & Serabyn, E. 2001, IEEE Transactions on Antennas and Propagation, 49, 1683

Pordes, R., Petravick, D., Kramer, B., et al. 2007, in 78, Vol. 78, J. Phys. Conf. Ser., 012057

Porth, O., Chatterjee, K., Narayan, R., et al. 2019, ApJS, 243, 26

Porth, O., Olivares, H., Mizuno, Y., et al. 2017, Computational Astrophysics and Cosmology, 4, 1

Prather, B., Wong, G., Dhruv, V., et al. 2021, The Journal of Open Source Software, 6, 3336

Prather, B. S., Dexter, J., Moscibrodzka, M., et al. 2023, ApJ, 950, 35

Psaltis, D., Johnson, M., Narayan, R., et al. 2018, arXiv e-prints, arXiv:1805.01242

Psaltis, D., Medeiros, L., Christian, P., et al. 2020, Phys. Rev. Lett., 125, 141104

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, ApJ, 885, 131

Röder, J., Cruz-Osorio, A., Fromm, C. M., et al. 2022, in European VLBI Network Mini-Symposium and Users' Meeting 2021, 24

Röder, J., Cruz-Osorio, A., Fromm, C. M., et al. 2023, A&A, 671, A143

Roelofs, F., Blackburn, L., Lindahl, G., et al. 2023, Galaxies, 11, 12

Roelofs, F., Fromm, C. M., Mizuno, Y., et al. 2021, A&A, 650, A56

Roelofs, F., Janssen, M., Natarajan, I., et al. 2020, A&A, 636, A5

Satapathy, K., Psaltis, D., Özel, F., et al. 2022, ApJ, 925, 13

Schmidt, M. 1963, Nature, 197, 1040

Sfiligoi, I., Bradley, D. C., Holzman, B., et al. 2009, in 2, Vol. 2, 2009 WRI World Congress on Computer Science and Information Engineering, 428–432

Simon, D. A., Cappellari, M., & Hartke, J. 2024, MNRAS, 527, 2341

Thompson, A. R., Moran, J. M., & Swenson, George W., J. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition (Springer)

van Bemmel, I. M., Kettenis, M., Small, D., et al. 2022, PASP, 134, 114502

van Langevelde, H. J., Frail, D. A., Cordes, J. M., & Diamond, P. J. 1992, ApJ, 396, 686

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

Wei, S.-W. & Liu, Y.-X. 2013, J. Cosmology Astropart. Phys., 2013, 063

Wielgus, M., Lančová, D., Straub, O., et al. 2022a, MNRAS, 514, 780

Wielgus, M., Marchili, N., Martí-Vidal, I., et al. 2022b, ApJ, 930, L19

Wong, G. N., Prather, B. S., Dhruv, V., et al. 2022, ApJS, 259, 64

Yao-Yu Lin, J., Pesce, D. W., Wong, G. N., et al. 2021, arXiv e-prints, arXiv:2110.07185

Younsi, Z., Porth, O., Mizuno, Y., Fromm, C. M., & Olivares, H. 2020, in Perseus in Sicily: From Black Hole to Cluster Outskirts, ed. K. Asada, E. de Gouveia Dal Pino, M. Giroletti, H. Nagai, & R. Nemmen, Vol. 342, 9–12

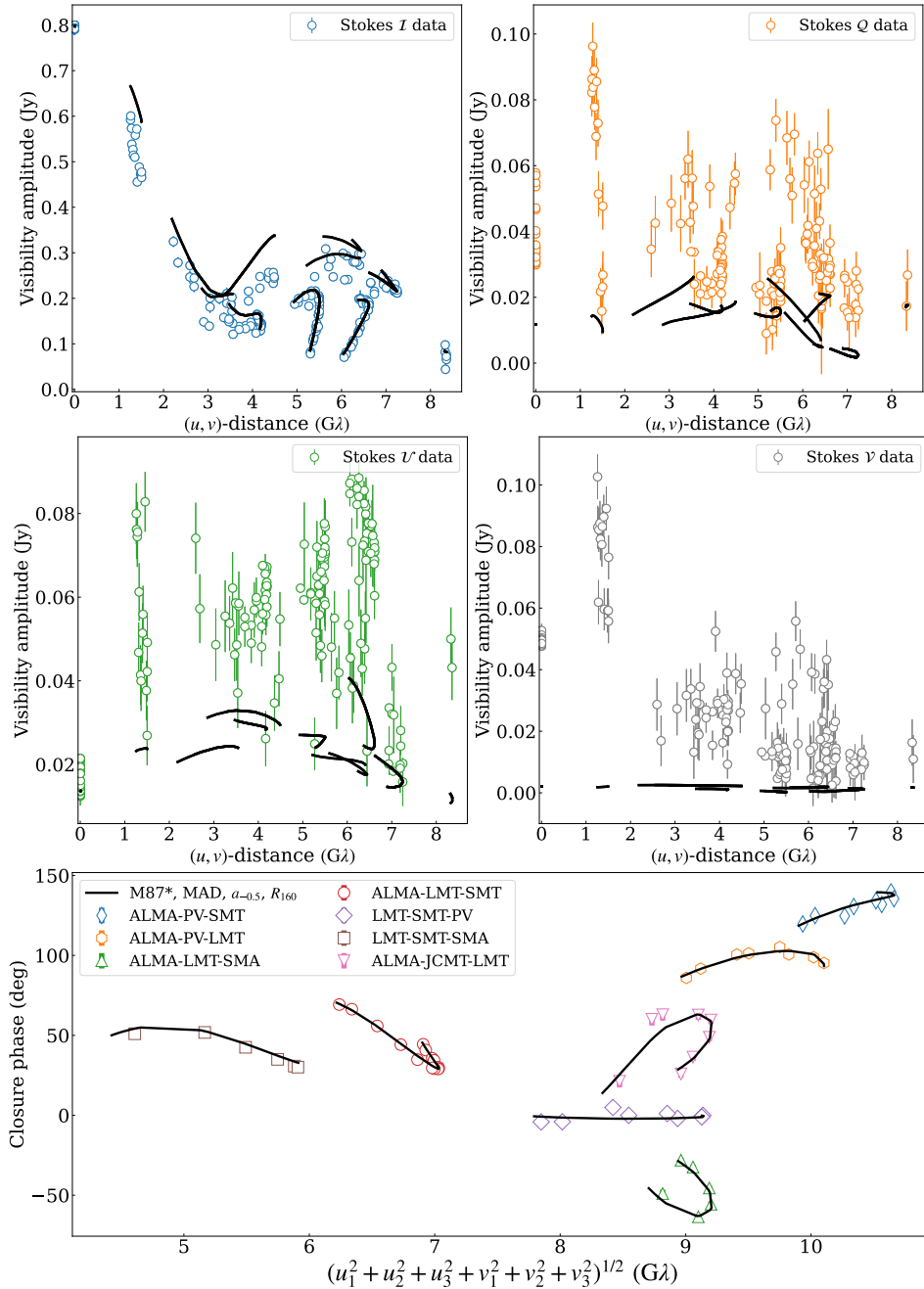## Appendix A: Synthetic data versus ground truth

Fig. A.1. Direct comparison of synthetic- and corresponding model data from the M87* MAD image of Figure 3. The four top panels show amplitudes of the different Stokes parameters being primarily affected by RCP, LCP gain errors and $\mathcal{D}$-terms. The closure phases shown in the bottom panel are mostly unaffected by the errors along the VLBI signal path. The synthetic data are averaged over VLBI scan durations. For most Stokes $\mathcal{I}$ and closure phase data points, the displayed thermal noise error bars are smaller than the plotted symbols. The model data is shown with a black line.

In Figure A.1, we show how the various simulated errors along the signal path affect the measured data, compared to model data produced by just an FFT as it would be measured by a perfect instrument. The Stokes $\mathcal{I}$ data is primarily affected by the the gross telescope gain errors $\mathcal{G}_{\mathrm{err}}$ that affect both polarizations in the same way, as well as antenna pointing errors. These errors are responsible for a loss of measured signal amplitude on several baselines. Due to polarization leakage, artificial source polarization is created, causing the linear polarization synthetic data amplitude to exceed the model. Circular polarization is very weak in the source and the produced signal is almost entirely the result of R−L telescope gain errors. The only baseline-dependent effect simulated by SYMBA is phase decoherence from atmospheric fluctuations. As those are well calibrated by RPICARD, the simulated closure phases show no substantial deviations from the model. The analysis of the Sgr A* synthetic data output gives similar results, albeit with additional noise due to the intrinsic source variability and the interstellar scattering screen.