

Decomposing the internal faraday rotation of black hole accretion flows

Angelo Ricarte,^{1,2}★ Ben S. Prather,³ George N. Wong,³ Ramesh Narayan,^{1,2} Charles Gammie^{3,4}
and Michael D. Johnson^{1,2}

¹Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

²Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA

³Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801, USA

⁴Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA

Accepted 2020 September 1. Received 2020 September 1; in original form 2020 July 21

ABSTRACT

Faraday rotation has been seen at millimeter wavelengths in several low-luminosity active galactic nuclei, including Event Horizon Telescope (EHT) targets M87* and Sgr A*. The observed rotation measure (RM) probes the density, magnetic field, and temperature of material integrated along the line of sight. To better understand how accretion disc conditions are reflected in the RM, we perform polarized radiative transfer calculations using a set of general relativistic magnetohydrodynamic (GRMHD) simulations appropriate for M87*. We find that in spatially resolved millimetre wavelength images on event horizon scales, the RM can vary by orders of magnitude and even flip sign. The observational consequences of this spatial structure include significant time-variability, sign-flips, and non- λ^2 evolution of the polarization plane. For some models, we find that internal RM can cause significant bandwidth depolarization even across the relatively narrow fractional bandwidths observed by the EHT. We decompose the linearly polarized emission in these models based on their RM and find that emission in front of the mid-plane can exhibit orders of magnitude less Faraday rotation than emission originating from behind the mid-plane or within the photon ring. We confirm that the spatially unresolved (i.e. image integrated) RM is a poor predictor of the accretion rate, with substantial scatter stemming from time variability and inclination effects. Models can be constrained with repeated observations to characterize time variability and the degree of non- λ^2 evolution of the polarization plane.

Key words: accretion, accretion discs – black hole physics – polarization – techniques: polarimetric – galaxies: individual (M87) – magnetohydrodynamics (MHD).

1 INTRODUCTION

Using a network of millimetre telescopes around the world, the Event Horizon Telescope (EHT) has recently produced the first images of a black hole (BH) accretion flow (Event Horizon Telescope Collaboration 2019a,b,c,d,e,f). These images resolve the ‘shadow’ of the supermassive BH M87*, corresponding to rays that begin on its event horizon, providing new constraints on the properties of the BH and its accretion disc.

In Event Horizon Telescope Collaboration (2019e), henceforth EHT5, a library of general relativistic magnetohydrodynamic (GRMHD) simulations was produced to compare to the EHT images, exploring three fundamental quantities. The first is the strength of the magnetic field: models that accumulate strong magnetic flux around the BH are able to counteract the ram pressure of in-falling gas with magnetic pressure, resulting in what is termed a magnetically arrested disc (MAD; Igumenshchev, Narayan & Abramowicz 2003; Narayan, Igumenshchev & Abramowicz 2003; Chael, Narayan & Johnson 2019). In contrast, the weaker and more turbulent magnetic fields in Standard And Normal Evolution (SANE) models have a weaker effect on the gas dynamics of the disc (Narayan et al.

2012; Sądowski et al. 2013; Ryan et al. 2018). The second quantity is the BH’s angular momentum, described by the dimensionless spin parameter $a \equiv Jc/(GM^2) \in (-1, 1)$, where negative values correspond to counter-rotating accretion discs. Mass and spin are the only properties intrinsic to an astrophysical BH, but BH spins are constrained much more loosely than their masses. Most of our understanding of supermassive BH spin evolution originates from theory (e.g. Bardeen & Wagoner 1969; Thorne 1974; Gammie, Shapiro & McKinney 2004; King, Pringle & Hofmann 2008; Volonteri et al. 2013). The third quantity explored in this work is R_{high} , one parametrization of the relative temperatures of electrons and ions in the plasma. Such a prescription is necessary because the mean free path near the event horizon is so large, the two populations depart from thermal equilibrium and divide into a two-temperature plasma (Shapiro, Lightman & Eardley 1976; Rees et al. 1982; Narayan & Yi 1995a; Sądowski et al. 2017; Ryan et al. 2018). By combining EHT imaging with other multiwavelength constraints such as the X-ray flux and jet power, EHT5 rule out all but 19/60 possible models that span these three variables. In particular, all $a = 0$ models are excluded.

EHT constraints on M87* thus far have only considered total intensity (Stokes I), when in fact polarized visibilities have been obtained (Stokes I , Q , U , and V), but have not been published. Hence, current constraints have only utilized one fourth of the

★ E-mail: angelo.ricarte@cfa.harvard.edu

measured information. Synchrotron emission, the emission mechanism at millimetre wavelengths, has a near-unity intrinsic polarization fraction (Le Roux 1961; Bromley, Melia & Liu 2001; Broderick & Blandford 2003). Polarimetric imaging is consequently predicted to tightly constrain accretion models, which differ substantially in their linear polarization fractions as well as their morphologies (Palumbo, Wong & Prather 2020). Although an image has not yet been constructed for Sgr A*, strong and variable linear polarization has been observed for this source. Previous very long baseline interferometry measurements have revealed a partially ordered magnetic field structure at its centre (Johnson et al. 2015).

As polarized radiation travels through a magnetized plasma, it is transformed by effects sensitive to the local density, temperature, and magnetic field. Faraday rotation turns the electric vector position angle (EVPA) of linearly polarized emission, and Faraday conversion exchanges linearly and circularly polarized radiation (Sazonov 1969; Rybicki & Lightman 1986; Melrose 1997; Shcherbakov 2008). Future analyses with the EHT will be able to further distinguish models in the time and frequency domains (e.g. Broderick & Loeb 2006a,b; Roelofs et al. 2017; Medeiros et al. 2018). It is therefore timely for us to better understand time and frequency-dependent effects that may help us constrain accretion models.

One such polarimetric observable is the rotation measure (RM), defined by the change in the EVPA, χ , as a function of the change of wavelength squared:

$$\text{RM} = \frac{\chi_2 - \chi_1}{\lambda_2^2 - \lambda_1^2}, \quad (1)$$

where the subscripts 1 and 2 denote two different wavelengths. RM probes Faraday rotation, and has been used in a variety of contexts to infer magnetic field properties (e.g. Zavala & Taylor 2004; Brentjens & de Bruyn 2005; Frick et al. 2011; Agudo et al. 2018; Pasetto et al. 2018). For a source of polarized emission that is entirely behind a Faraday rotating medium, the RM can be written as an integral of plasma properties along the line-of-sight via

$$\text{RM} = 8.1 \times 10^5 \text{ rad m}^{-2} \int_{\text{source}}^{\text{observer}} f_{\text{rel}}(\Theta_e) \frac{n_e}{1 \text{ cm}^{-3}} \frac{B_{\parallel}}{G} \frac{ds}{\text{pc}}, \quad (2)$$

where n_e is the electron number density, B_{\parallel} is the parallel component of the magnetic field, and f_{rel} is a correction term suppressing Faraday rotation at relativistic temperatures (Gardner & Whiteoak 1966). At relativistically hot temperatures, $f_{\text{rel}}(\Theta_e) \approx \log(\Theta_e)/(2\Theta_e^2)$, whereas at sub-relativistic temperatures, f_{rel} asymptotes to 1. Here, $\Theta_e \equiv k_B T_e / m_e c^2$, k_B is the Boltzmann constant, T_e is the electron temperature, m_e is the electron rest mass, and c is the speed of light (Jones & Odell 1977).

As seen from equation (2), the RM directly traces the electron temperature, number density, and magnetic field along the line of sight. Among the models consistent with the EHT data of M87*, all of these quantities can vary by orders of magnitude. Note that while RM scales as $\text{RM} \propto n_e B$, the power emitted by synchrotron emission scales as $P \propto n_e B^2$ (Rybicki & Lightman 1986). In principle, this could allow RM to break a degeneracy that exists between n_e and B based on total intensity alone.

RMs have been measured for the two main EHT targets Sgr A* and M87*, as well as a handful of other low-luminosity AGN. For Sgr A*, $\text{RM} = -5 \times 10^5 \text{ rad m}^{-2}$ (Bower et al. 2003; Marrone et al. 2007; Bower et al. 2018), while $-7 \times 10^4 \text{ rad m}^{-2}$ has been measured a few arcseconds away (Eatough et al. 2013). For M87*, an upper limit of $|\text{RM}| < 7 \times 10^5 \text{ rad m}^{-2}$ was measured using the Submillimeter Array (Kuo et al. 2014). 3C 84 has an RM of $10^{5-6} \text{ rad m}^{-2}$ (Kim et al. 2019) at 43 GHz and $9 \times 10^5 \text{ rad m}^{-2}$ at 230 GHz (Plambeck

et al. 2014). Similarly, 3C 273 has an RM of $5 \times 10^5 \text{ rad m}^{-2}$ at 230 GHz (Hovatta et al. 2019). Neither linear polarization nor RM could be measured for the low-luminosity AGNs M81 and M84, which might imply significant scrambling (Brunthaler et al. 2001; Bower et al. 2017).

These measurements have been used to constrain the accretion rates of EHT targets Sgr A* and M87* by assuming simple analytic models describing the accretion flow that can be input into equation (2; Marrone et al. 2006). The accretion rates of Sgr A* and M87* have thus been constrained to $< 10^{-6} M_{\odot} \text{ yr}$ (although it may be much lower; Agol 2000; Quataert & Gruzinov 2000b; Marrone et al. 2006) and $< 9.2 \times 10^{-4} M_{\odot} \text{ yr}$ (Kuo et al. 2014; Li, Yuan & Xie 2016), respectively. However, there are important effects that complicate the simple scenario implicitly assumed by equation (2), where a Faraday rotator sits entirely between a source and our line of sight (see also Broderick & McKinney 2010). Most importantly, for BH accretion flows, Faraday rotation and emission occur co-spatially, such that along a given geodesic, not all photons are Faraday rotated by the same material. In addition, general relativity (GR) complicates a photon's trajectory and can modify its polarization properties by parallel transport alone. This is especially true of emission near the photon ring, where null geodesics can pass through the accretion flow multiple times in different directions, leading to interesting polarization signatures (Johnson et al. 2019; Himwich et al. 2020). Finally, neither the emission nor the Faraday rotation can be assumed to behave in a spatially uniform way, and the magnetic field may switch sign, especially in turbulent SANE discs.

Mościbrodzka et al. (2017) produced the first polarized model images of M87* at millimetre wavelengths from ray traced GRMHD simulations. Using a set of SANE models, they determined that Faraday rotation can be strong enough to spatially scramble the polarization from the counter-jet, which must pass through a larger Faraday depth than the forward-jet on the way to the observer. Their analysis also revealed a significant inclination dependence of the RM, such that even very high accretion rate models could satisfy RM constraints when viewed face-on. Jiménez-Rosales & Dexter (2018) determined that high-accretion rate models are disfavoured, since this scrambling too strongly depolarizes the emission. Tsunetoe et al. (2020) began to explore the spin dependence and favoured $a = 0.9$ models, although an expanded parameter survey is warranted.

In this work, we perform a more comprehensive survey of RM for seven models consistent with EHT5, chosen to bracket the allowed parameter space. We consider variations as a function of time and inclination, and develop new techniques to model Faraday effects in some models. In Section 2, we describe the GRMHD simulations we use as a starting point, the radiative transfer calculations in post-processing, and a novel Taylor expansion model for treating internal Faraday rotation. In Section 3, we describe our results. This includes an exploration of strong spatial variations in RM, RM distribution functions, RM as a measure of accretion rate, inclination dependence, the degree of non- λ^2 behaviour among models, and a case study of time variability. We discuss our results in Section 4, and end with a summary and conclusion in Section 5.

2 METHODOLOGY

We begin with a set of GRMHD simulations that are consistent with EHT5. We then use IPOLE¹ (Mościbrodzka & Gammie 2018) to perform polarized radiative transfer calculations, specifying electron

¹<https://github.com/moscibrodzka/ipole/>

Table 1. Parameters of the seven models considered in this paper. Each of these models passes all metrics considered in EHT5 and are chosen to bracket the allowed parameter space.

Magnetic field state	a	R_{high}
MAD	+0.94	160
MAD	+0.94	20
MAD	−0.5	160
MAD	−0.5	20
SANE	+0.94	160
SANE	−0.94	80
SANE	−0.94	10

properties in this step. Finally, using a first-order Taylor expansion, we create a model of the polarized image in order to capture frequency-dependent effects and compute RMs.

2.1 GRMHD Simulations and Radiative Transfer

In this work, we study seven models for M87* in the EHT GRMHD simulation library, each of which is consistent with all of the observational constraints considered in EHT5. The properties of these simulations, all performed with IHARM (Gammie, McKinney & Tóth 2003), are listed in Table 1. Both SANE and MAD models are considered, while values of a and R_{high} are chosen to bracket the allowed values in EHT5. These models are described in more detail in EHT5, and the SANE $a = +0.94$ model is included in the GRMHD code comparison project of Porth et al. (2019).

As defined by Mościbrodzka, Falcke & Shiokawa (2016), R_{high} prescribes the electron temperature via

$$\frac{T_i}{T_e} = R_{\text{high}} \frac{\beta_p^2}{1 + \beta_p^2} + \frac{1}{1 + \beta_p^2}, \quad (3)$$

where T_i and T_e are the ion and electron temperatures, respectively, and β_p is the ratio of gas to magnetic pressure. T_i is determined by the GRMHD simulation. In the mid-plane, where β_p is high, $T_e \rightarrow T_i/R_{\text{high}}$, since turbulent plasma models reveal that heating preferentially affects ions, which then cannot efficiently transfer energy to electrons (Rees et al. 1982; Narayan & Yi 1995b; Quataert & Gruzinov 1999; Howes 2010; Kawazura, Barnes & Schekochihin 2019). Consequently, increasing R_{high} has the effect of decreasing the emission from and increasing the Faraday rotation within the mid-plane. As a result of decreasing mid-plane emission, the accretion rate must also be scaled upwards to obtain the correct total intensity for M87*. As we shall show, both of these effects have important implications for the RM.

Note that among these seven models, there are only four unique GRMHD simulations, since R_{high} only affects the radiative transfer in post-processing. In each of these simulations, the angular momentum of the disc is either perfectly aligned (denoted by positive spin) or antialigned (denoted by negative spin) with that of the BH, although misaligned discs remain an active area of research (e.g. Fragile et al. 2007; Chatterjee et al. 2020; Liska et al. 2020). MAD simulations are run with a $384 \times 192 \times 192$ grid with a maximum radius of $10^3 GM_{\bullet}/c^2$, while SANE simulations are run with a $288 \times 128 \times 128$ grid and a maximum radius of $50 GM_{\bullet}/c^2$. However, we find that these models only exhibit inflow equilibrium within a radius of approximately $20 GM_{\bullet}/c^2$. Throughout this work, we restrict the integration of our radiative transfer equations to within this radius. Fortunately, this limitation actually has negligible effect on our results for face-on inclinations ($i \lesssim 30^\circ$), as we explore in Appendix C. This is

because the funnel region is largely evacuated in these simulations and does not contribute to Faraday rotation. However, restricting our calculations to $\leq 20 GM_{\bullet}/c^2$ may lead us to underestimate the total Faraday rotation at inclinations $\gtrsim 30^\circ$, due to material that may exist in more distant, unequilibrated regions of the simulations.

We create polarized ray-traced images using IPOLE (Mościbrodzka & Gammie 2018), which first solves the null geodesic equation backwards from the image plane, then integrates forward the radiative transfer equations for the four Stokes parameters, $\{I, Q, U, V\}$. Here, I is the total intensity, $\sqrt{Q^2 + U^2}$ and $\frac{1}{2} \arg(Q + iU)$ are the linearly polarized intensity and EVPA, respectively, and V is the circularly polarized intensity. The radiative transfer equations account for synchrotron emission, synchrotron self-absorption, Faraday rotation, and Faraday conversion. Radiative transfer coefficients follow Dexter (2016) for a thermal electron distribution function, with a slight modification to $\rho_{v,v}$, the coefficient responsible for Faraday rotation. As also discussed in Dexter et al. (2020), minor modifications are needed to ensure continuous and accurate behaviour at low temperature and frequency. Following Shcherbakov (2008), we set

$$\rho_{v,v} = \frac{2ne^2\nu_B}{m_e c \nu^2} \frac{K_0(\Theta_e^{-1})}{K_2(\Theta_e^{-1})} \cos(\Theta_e) g(X), \quad (4)$$

where K_0 and K_2 are modified Bessel functions of the second kind, e is the electron charge, m_e is the electron mass, $\nu_B = eB/2\pi m_e c$, $X = \left[\frac{3}{2\sqrt{2}} 10^{-3} \frac{\nu}{\nu_c} \right]^{-1/2}$, and $g(X) = 1 - 0.11 \ln(1 + 0.035X)$ for cyclotron frequency ν_c .

For each model, we create images for 11 snapshots spanning the last quarter of the corresponding GRMHD run, corresponding to times $t/(GM_{\bullet}/c^3) \in [7500, 10000]$, a duration of 880 d for M87*. For each snapshot, we study five inclinations, $i \in \{5^\circ, 17^\circ, 30^\circ, 60^\circ, 90^\circ\}$.² Then, for each snapshot and inclination, we create six polarized images. We sample three frequencies, 226.999, 227.000, and 227.001 GHz, to construct a frequency-dependent model of the image, detailed in the following section. We find that we must adopt extremely small differences in frequency in order to resolve the RM on geodesics that have high Faraday depth. Otherwise, the RM in some geodesics can be underestimated due to the ‘ $n\pi$ degeneracy:’ the EVPA may rotate so rapidly within that multiple rotations are missed. Then, for each of these three frequencies, we create two separate images that only include emission from either the positive or negative z domains, where the z -axis is oriented parallel to the BH spin (and perpendicular to the disc in these models).

Separate images including only the top and bottom halves of the emission are helpful for modelling Faraday effects that are often significantly different between the front and back sides of the emitting region. In these models, emission behind the mid-plane often experiences orders of magnitude more Faraday rotation than emission anterior to it, since it must pass through the relatively cold mid-plane (Mościbrodzka et al. 2017). This split remains helpful at high inclinations, since some emission is lensed to the opposite side of the image. A complete image is made by summing together the individual images made with only the near- and far-side emission. Note that when emission from only one side is included, absorption and Faraday effects from both sides remain included.

For all images, we adopt a BH mass of $6.2 \times 10^9 M_{\odot}$ and a distance of 16.9 Mpc (Gebhardt et al. 2011), for consistency with

²Throughout this paper, for models with positive spin, we actually compute images for $i = 180^\circ - i_{\text{written}}$, as in EHT5.

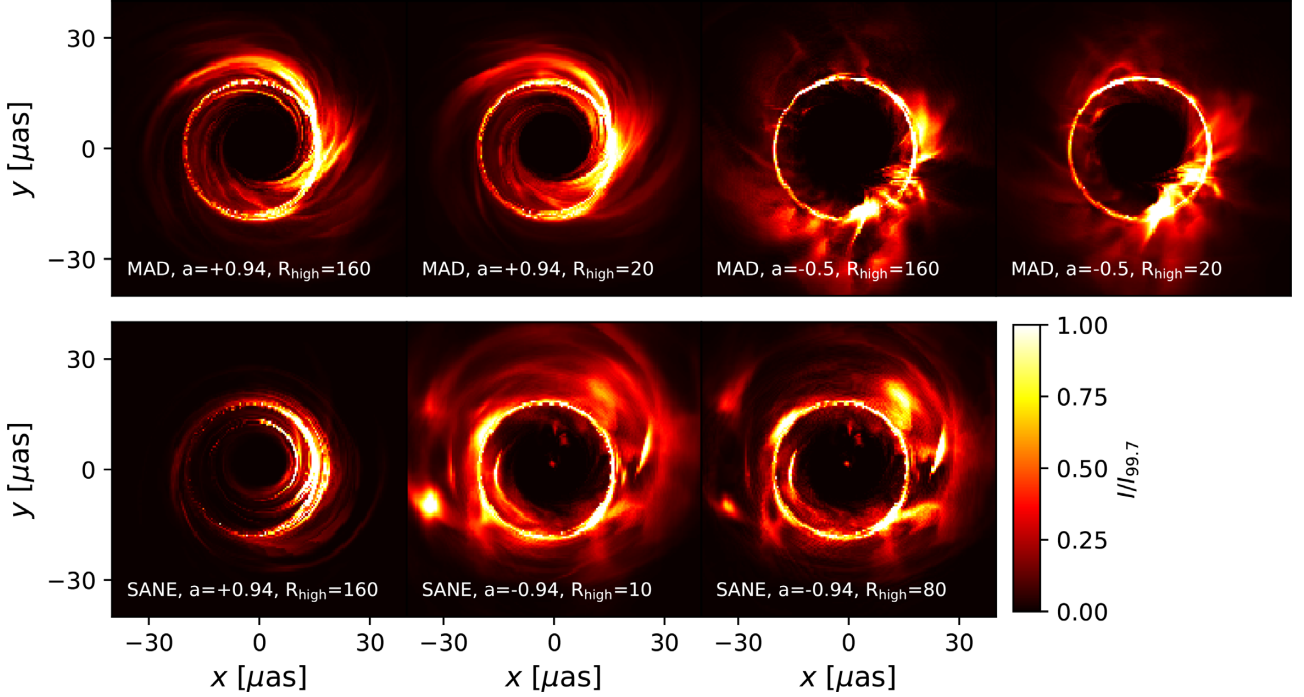


Figure 1. Total intensity images of the seven models considered in this work, taken at the final snapshot of each GRMHD simulation, which occurs at time $t = 10000 GM_*/c^3$. Models are tilted by 17° from face-on towards the top of these images and in subsequent figures. To help visualize low surface brightness features, intensity is scaled with respect to the intensity of the pixel in the 99.7th percentile, $I_{99.7}$, saturating 0.3 per cent of the pixels.

EHT5. Gas densities are scaled such that the average image produces an intensity of 0.5 Jy at an inclination of 17° , which is the most likely inclination at which we are viewing M87* based on its larger scale jet (Walker et al. 2018). The same gas density scalings are used for models at different inclinations, although their average fluxes can depart from 0.5 Jy. We find that the total intensities of images created at an inclination of 90° tend to be approximately a factor of 2 larger than those created with an inclination of 17° . Each image is created at $0.5 \mu\text{as}$ pixel resolution, with a $160 \mu\text{as}$ field of view, a factor of 2 finer angular resolution than employed in EHT5.

To summarize, we consider seven models that are consistent with EHT observations of M87*. For each model, we create images for 11 snapshots, 5 inclinations, 3 frequencies, and the 2 sides of the accretion flow. We also create additional images to better resolve frequency-dependent (Section 2.2) and time-dependent (Section 3.7) phenomena. In Fig. 1, we plot total intensity images of the seven models during their final snapshot at 17° inclination. To better visualize low surface brightness features, we intentionally saturate 0.3 per cent of the pixels in this and all subsequent visualizations. In all of the images presented in this work, the forward-jet points straight up, and the material on the right side of the image is moving towards the observer.

2.2 Modelling the polarized image

Here, we introduce a modelling formalism for the polarized image as a function of frequency. For a given image snapshot, we create a two-zone (front and back half) model of the radiation in each pixel. Our model is constructed as follows:

(i) Treating the front and back halves of the emitting region separately, we first convert the Stokes parameters $\{I, Q, U, V\}$ into

their counterparts on the Poincare sphere, $\{I, N, \phi, \psi\}$ (following the notation of Shcherbakov, Penna & McKinney 2012). These variables have well-behaved Taylor series expansions, and are further discussed in Appendix A.

(ii) Using images at three frequencies, we then compute the first and second derivatives of the spherical Stokes parameters. Each derivative is computed in the variable that best approximates expected physical dependencies: I and N are differentiated in ν , ϕ is differentiated in λ^2 , and ψ is differentiated in λ^3 .

(iii) Using the first derivatives, we construct Taylor expansions to first order in each of the spherical Stokes parameters. These can be expressed

$$I(\nu) \approx I(\nu_0) + \frac{dI}{d\nu}(\nu - \nu_0), \quad (5)$$

$$N(\nu) \approx N(\nu_0) + \frac{dN}{d\nu}(\nu - \nu_0), \quad (6)$$

$$\phi(\nu) \approx \phi(\nu_0) + \frac{d\phi}{d\lambda^2}(\lambda^2 - \lambda_0^2) \approx \phi(\nu_0) - \frac{2c^2}{\nu_0^3} \frac{d\phi}{d\lambda^2}(\nu - \nu_0), \quad (7)$$

$$\psi(\nu) \approx \psi(\nu_0) + \frac{d\psi}{d\lambda^3}(\lambda^3 - \lambda_0^3) \approx \psi(\nu_0) - \frac{3c^3}{\nu_0^4} \frac{d\psi}{d\lambda^3}(\nu - \nu_0). \quad (8)$$

(iv) In order to approximate bandwidth integration, we integrate analytically the equations for the Stokes parameters $\{I, Q, U, V\}$. For a band spanning the frequencies ν_1 and ν_2 , the bandwidth-averaged Stokes parameters are given by

$$I_{\text{BW}} = \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} I(\nu) d\nu, \quad (9)$$

$$Q_{\text{BW}} = \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} N(\nu) \cos[\phi(\nu)] \sin[\psi(\nu)] d\nu, \quad (10)$$

$$U_{\text{BW}} = \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} N(\nu) \sin[\phi(\nu)] \sin[\psi(\nu)] d\nu, \quad (11)$$

$$V_{\text{BW}} = \frac{1}{\nu_2 - \nu_1} \int_{\nu_1}^{\nu_2} N(\nu) \cos[\psi(\nu)] d\nu. \quad (12)$$

Analytic solutions to these integrals are provided in Appendix B.

(v) Finally, the front and back emission halves are summed to produce the complete image.

In practice, $d\phi/d\lambda^2$, which encapsulates Faraday rotation, is the most important frequency-dependent effect to take into account. Faraday conversion, which is encapsulated in $d\psi/d\lambda^3$, is not as strong as Faraday rotation in these models. In some of our models, our split into the front and back halves is necessary for capturing the Faraday depolarization of back half of the emission (due to the Faraday thick mid-plane) while preserving the polarization of the front half. As we will later explore in more detail, some pixels may simultaneously have emission from the forward-jet with an RM of $\sim 10^3 \text{ rad m}^{-2}$, and emission from the counter-jet with an RM of $\sim 10^9 \text{ rad m}^{-2}$. Consequently, $\phi(\lambda^2)$ for the pixel is highly non-linear, but can be approximated as the sum of two emitting regions with distinct RMs.

In some pixels, especially within the photon ring, we compute unreasonably high values of $|dN/d\nu|$ within our Taylor expansion. This is due to multiple emission components in the same pixel Faraday rotating at different rates, which can periodically increase and decrease the total linearly polarized intensity in the pixel depending on the relative phases of these components. We use $\frac{d^2N}{d\nu^2}$ to help identify such problematic pixels. Let $\Delta\nu = \nu - \nu_0$ be the difference in frequency space between some frequency ν at which Stokes parameters are being evaluated and the frequency at which derivatives have been computed, $\nu_0 = 228 \text{ GHz}$. Treating $\frac{1}{2} \frac{d^2N}{d\nu^2} \Delta\nu^2$ as an upper limit on the error of $N(\nu)$, we freeze the value of N in pixels where $|\frac{dN}{d\nu} \Delta\nu| < |\frac{1}{2} \frac{d^2N}{d\nu^2} \Delta\nu^2|$ and $|\frac{d \ln(N)}{d\nu}| > |\frac{1}{\Delta\nu}|$. That is, we freeze N if its first derivative implies that it would grow or shrink by more than a factor of e across $\Delta\nu$ and the absolute error on the change of N may be larger than the change of N itself. We apply an identical condition on the derivative of I , which almost always affects pixels also affected by the condition on Stokes N .

We examine to what extent this criterion is applied in our final snapshot images, and find that it affects only a minority of pixels. Among the pixels which amount to 99 per cent of the image integrated I , we find that at most 5 per cent of the pixels are affected by this correction, typically in the photon ring. This largest fraction occurs in the SANE, $a = -0.94$, $R_{\text{high}} = 80$ model, which as we shall show contains the emission with the largest Faraday depths. In some of the other models, none of the pixels are affected by this criterion.

In Fig. 2, we demonstrate the efficacy of our method by plotting the linearly polarized intensity ($\sqrt{Q^2 + U^2}$) for the SANE, $a = +0.94$, $R_{\text{high}} = 160$ model. The top left-hand panel shows the result of a single-frequency calculation at 228 GHz, while the top right image shows the result of averaging 255 images across a 4 GHz bandwidth between 226 and 230 GHz. In the bandwidth-averaged image, much of the large-scale emission has been suppressed and two streaks of emission in the upper right of the image have become more prominent. The total spatially unresolved linear polarization fraction has dropped by a factor of 2.4. Although the counter-jet dominates the linearly polarized emission in the single-frequency image due to lensing (Dexter, McKinney & Agol 2012), most of that emission is scrambled away due to the large Faraday depth in the mid-plane (Mościbrodzka et al. 2017).

The bottom left-hand panel is the result of our Taylor approximated model based on six images (three frequencies, two sides), which successfully captures the depolarization of the counter-jet, but not the forward-jet. In the bottom right-hand panel, we subtract this

model from the averaged images and plot the residual. Weighting by the linear polarized intensity of the properly averaged image, total linear polarization is recovered on average with an absolute error of $0.32 \mu\text{Jy}$ and a relative error of 0.081. In contrast, the single-frequency image has an average absolute error $5.7 \mu\text{Jy}$ and relative error of 4.0. In this case, the single-frequency image also overestimates the total linear polarization somewhat, even if the source is spatially unresolved. When Stokes parameters are summed across the entire image, the properly averaged image has a linear polarization fraction of 8.3×10^{-3} , the Taylor approximated image has a linear polarization fraction of 7.3×10^{-3} , and the single-frequency image has a linear polarization fraction of 1.9×10^{-2} . A model containing at least two separate zones is necessary in order to reproduce the depolarization of the counter-jet without also depolarizing the forward-jet. The remaining residuals require more than two regions to fully capture the complexity of the Faraday rotating structure along the line of sight. Notice that the most significant errors occur within the photon ring or its interior, where geodesics pass through the mid-plane multiple times.

In Fig. 3, we further decompose the single-frequency 228 GHz image into emission from its forward-jet and from counter-jet components. In the lower panel, we examine frame-invariant³ radiative transfer coefficients in the pixel marked by a blue circle in these images. In this pixel, polarized emission is emitted roughly equally by forward-jet and counter-jet components. However, the counter-jet emission must pass through the enormous Faraday depth of the cold mid-plane, with $\text{RM} > 10^9 \text{ rad m}^{-2}$ in some regions. The polarized intensity of emission passing through material with a rotation measure of RM and a bandwidth of $\Delta\nu$ is suppressed by a factor f_{BW} given by

$$f_{\text{BW}} = \text{sinc}\left(\frac{2c^2}{v_0^3} \text{RM} \Delta\nu\right), \quad (13)$$

where $\text{sinc}(x) = \sin(x)/x$, c is the speed of light, and v_0 is the central frequency, assuming narrow fractional bandwidth and uniform sampling in ν . We define the critical rotation measure RM_{crit} via

$$\frac{1}{2} = \text{sinc}\left(\frac{2c^2}{v_0^3} \text{RM}_{\text{crit}} \text{BW}\right). \quad (14)$$

That is, RM_{crit} is the minimum RM required to suppress linear polarization by a factor of 2. So far, EHT observations have been performed using a central frequency of 228 GHz and a bandwidth of 4 GHz. Using these EHT values, we find that $\text{RM}_{\text{crit}} = 3 \times 10^7 \text{ rad m}^{-2}$. Since $10^9 \text{ rad m}^{-2} \gg \text{RM}_{\text{crit}}$, the counter-jet's polarization is significantly suppressed. Notice that the photon ring emission from the forward-jet is also suppressed, since this emission must also pass through the cold mid-plane. Jiménez-Rosales & Dexter (2018) determine that strong Faraday effects also scramble the image on a pixel-by-pixel basis, resulting in beam depolarization. This spatial decoherence helps compensate for bandwidth depolarization when blurred images are constructed.

2.3 Rotation measure

In Section 2.2, we compute Stokes parameters and their derivatives at a central frequency of 228 GHz and a small bandwidth of 2 MHz.

³Since ν varies along the geodesic, it is useful to define the frame-invariant quantities $\rho\nu = \nu\rho_{\nu,\nu}$ and $j_Q = j_\nu/\nu^2$ (for additional discussion, see Mościbrodzka & Gammie 2018).

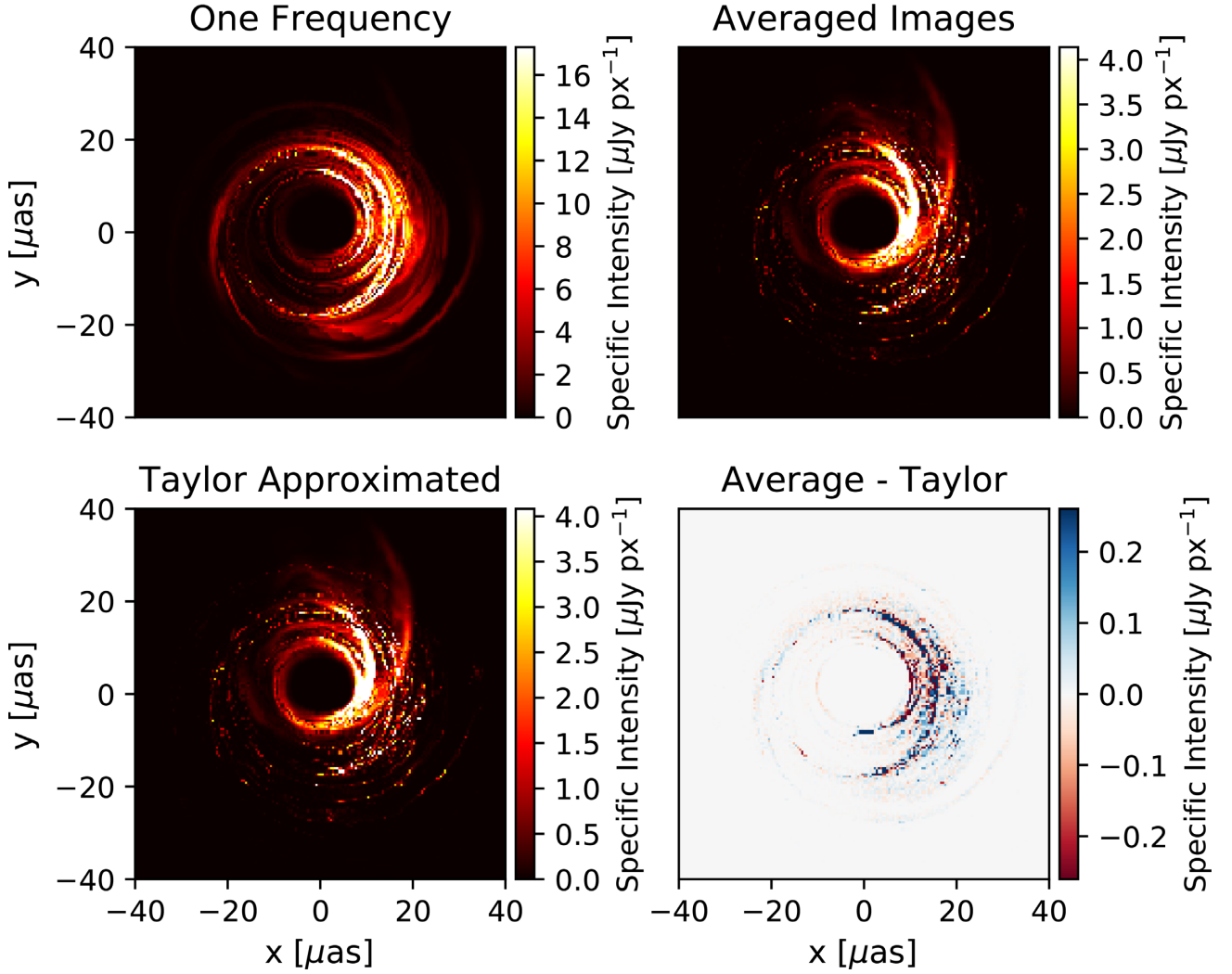


Figure 2. Images of total linearly polarized intensity ($\sqrt{Q^2 + U^2}$) for a case where bandwidth depolarization is important, the SANE, $a = +0.94$, $R_{\text{high}} = 160$ model, centred at 228 GHz. In the top left, we plot an image at a single frequency, while in the top right, we plot the average of 255 images across a 2 GHz bandwidth. As further explored in Fig. 3, bandwidth depolarization suppresses the contribution of the counter-jet. In the bottom left, we plot the result of our Taylor expansion model based on six images. Its residual with respect to the 255 averaged images is shown in the bottom right. The Taylor expansion model successfully depolarizes the counter-jet without depolarizing the forward-jet. Since we model the image with two zones, the largest discrepancies occur in and within the photon ring, where geodesics cross the mid-plane multiple times.

From these, we can directly compute the RM in each pixel at 228 GHz via

$$\text{RM} \equiv \frac{d\chi}{d\lambda^2} = \frac{d}{d\lambda^2} \frac{1}{2} \arctan\left(\frac{U}{Q}\right) = \frac{1}{2} \frac{U'Q - Q'U}{Q^2 + U^2}, \quad (15)$$

which follows from a straightforward application of the chain rule; here the $'$ symbols denote $d/d\lambda^2$. When we discuss the RM of individual pixels, this is how RM is computed. Using a small band of 2 MHz, the maximum measurable RM in an individual pixel is $|\text{RM}|_{\text{max}} = \pi/\Delta\lambda^2 \approx \pi\nu_0^3/(2c^2\Delta\nu) = 1.0 \times 10^{11} \text{ rad m}^{-2}$.

However, when discussing the RM for spatially unresolved measurements in subsequent sections, it is important to recognize the complicated evolution of $\chi(\lambda^2)$ that results from the complex RM structure we uncover in Section 3.1. When assigning a single value of the RM to an entire spatially unresolved image, we use the Taylor expansion methodology developed in Section 2.2 to approximate 16 polarized images, each integrated over a bandwidth of 0.25 GHz, equally spaced in λ^2 space between 226 and 230 GHz to emulate

EHT observations. These images are used to compute $\chi(\lambda^2)$, and $\Delta\chi$ is computed from the endpoints of the band. We correct for phase wraps by adding or subtracting π to $\chi(\lambda^2)$ as necessary to obtain the correct sign of $d\chi/d\lambda^2$ based on equation (15). Using 16 bands of width 0.25 GHz, the maximum measurable RM is $|\text{RM}|_{\text{max}} = \pi/\Delta\lambda^2 \approx \pi\nu_0^3/(2c^2\Delta\nu) = 8.3 \times 10^8 \text{ rad m}^{-2}$.

3 RESULTS

3.1 Structure of spatially resolved RM

In this section, we study the RM of each of our models within single simulation snapshots. We find significant spatial variation across the image due to inhomogeneities in the accretion flow on event horizon scales. At different locations, RM can vary by orders of magnitude and even flip sign. As a result, for spatially unresolved polarized measurements, some of these models may exhibit highly non-linear $\chi(\lambda^2)$, making them difficult to characterize with a single RM.

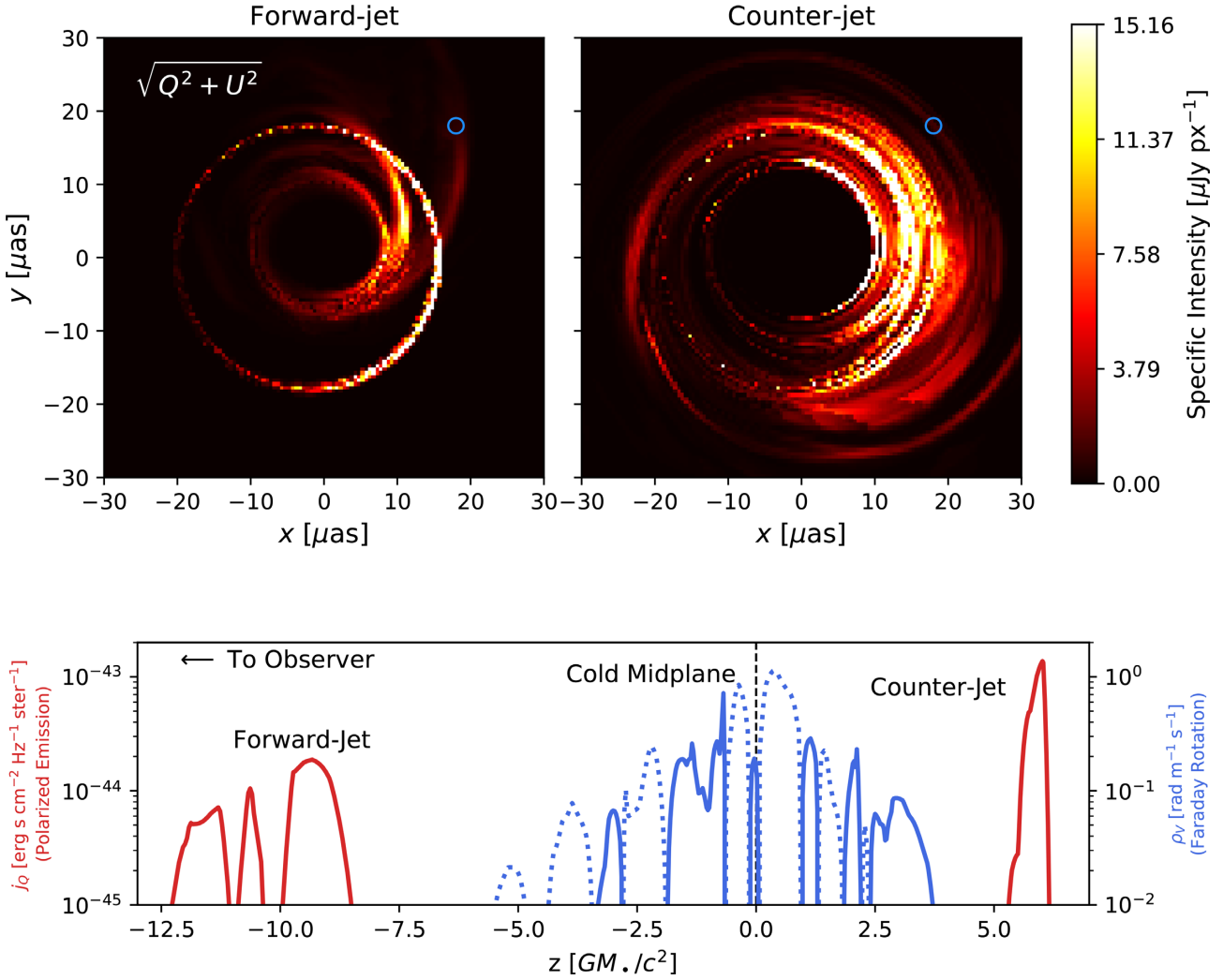


Figure 3. Linearly polarized intensity ($\sqrt{Q^2 + U^2}$) for the SANE, $a = +0.94$, $R_{\text{high}} = 160$ model at 227 GHz (0 bandwidth), broken into forward-jet (Top Left) and counter-jet (Top Right) emission components, which appear very different due to lensing. For the geodesic marked by a blue circle, we plot in the lower panel the frame-invariant versions of the radiative transfer coefficients that correspond to polarized emissivity (j_0 , red) and Faraday rotation (ρ_V , blue). For ρ_V , the dotted lines correspond to negative values, and there are many sign flips due to the turbulent nature of a SANE disc. Emission from the counter-jet passes through the cold mid-plane on the way to the observer, which produces a very large RM. Thus, when properly averaging over bandwidth, the counter-jet emission is suppressed by bandwidth depolarization. (See Fig. 2, top right-hand panel.).

In Fig. 4, we visualize the spatially resolved RM structure of the images presented in Fig. 1. In this figure, the brightness of each pixel scales with the linearly polarized specific intensity ($\sqrt{Q^2 + U^2}$), while the colouration encodes the RM. As in Fig. 1, 0.3 per cent of the pixels are saturated to more clearly display low surface brightness structures. The colour scale is normalized separately for each model, spanning \pm the 90th percentile of $|\text{RM}|$ of the pixels that have at least 50 per cent of the maximum linear polarized intensity plotted. The red regions have positive RM, while the blue regions have negative RM. The top two rows depict the RM with $\Delta\nu = 2$ MHz, where our Taylor expansion model is constructed. The bottom two rows plot the RM across a 4 GHz band, a more realistic bandwidth, within which bandwidth depolarization becomes important for some models. Images of 4 GHz are plotted with the same brightness and RM scale as their 2 MHz counterparts.

As shown in this figure, complex spatial variation and frequent sign-flips are a generic feature of these RM maps. This behaviour is not surprising in SANE models, which are characterized by weaker,

disordered magnetic fields, but is less expected in MAD models, which are characterized by strong poloidal fields. In one suggestive snapshot, we confirm that these RM sign-flips are due to sign-flips in the magnetic field with respect to the geodesic. Fig. 5 plots the intensity-weighted Faraday depth in each pixel, $\tau_F = \int \rho_V ds$, for a snapshot of the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model. Here, ρ_V is the (frame-invariant) radiative transfer coefficient responsible for Faraday rotation, and s is the affine parameter describing the geodesic. The sign of this quantity, shown to exhibit both positive to negative values, directly encodes the direction of the magnetic field with respect to the photon trajectory. RM sign flips have been predicted by earlier MHD simulations without GR, but only at large inclinations (Sharma et al. 2007).

By performing a 3D visualization of this snapshot with VISIT (Childs et al. 2012), we find that sign flips in the magnetic field occur in its tangential and radial components when crossing the disc mid-plane. Near the event horizon, field lines on the Northern hemisphere point west, while field lines on the Southern hemisphere point east,

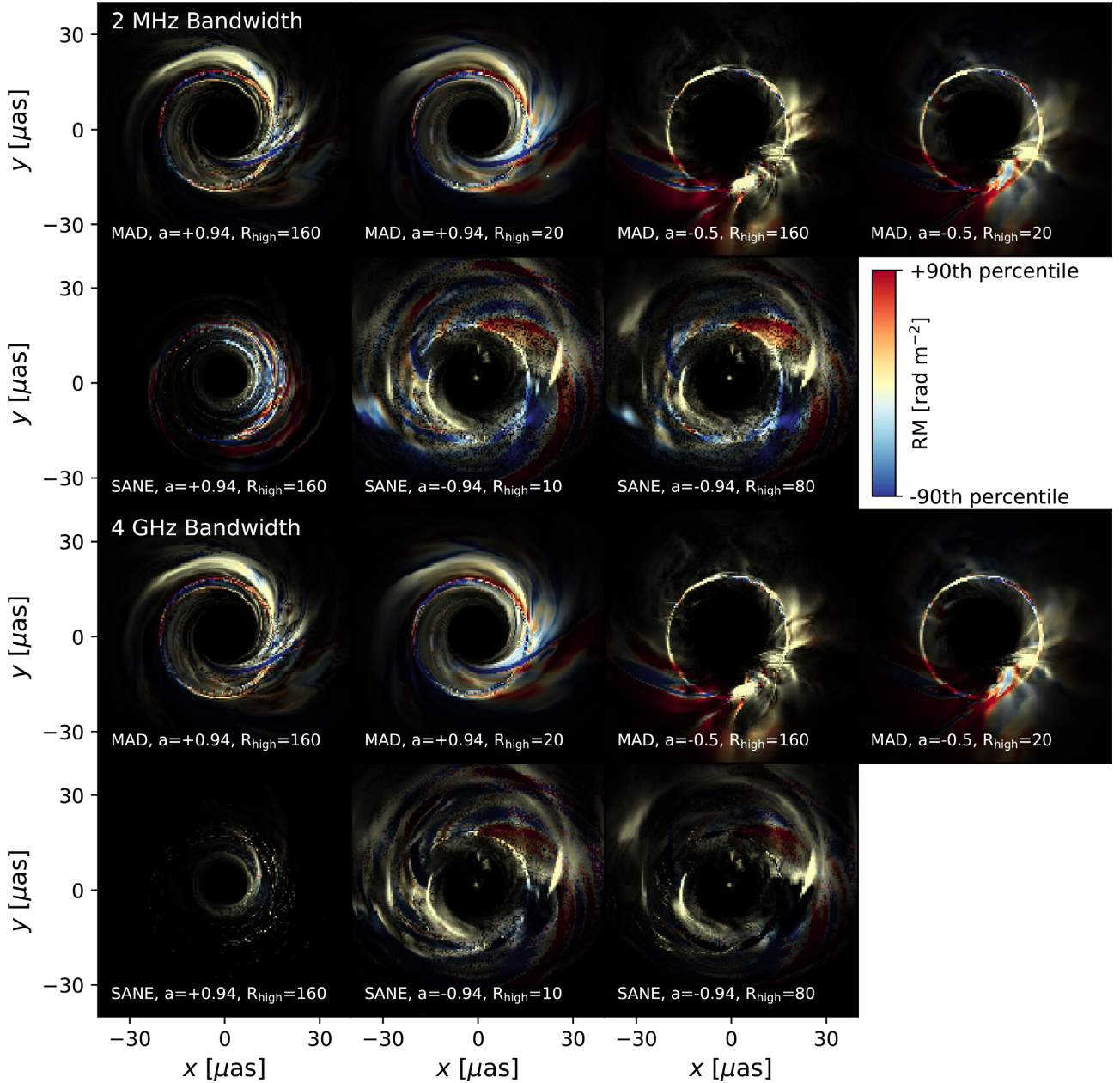


Figure 4. Visualization of the RM structure for the snapshots plotted in Fig. 1. The brightness of each pixel is proportional to the linearly polarized intensity in the pixel (again saturating 0.3 per cent of the pixels), while the colour encodes the RM. The colour scale is normalized separately for each model, spanning \pm the 90th percentile of $|RM|$ of the pixels which have at least 50 per cent of the maximum linear polarized intensity. The red regions have positive RM, while the blue regions have negative RM. The top two rows are not bandwidth corrected, while the bottom two are. This can dramatically affect the SANE models, but has little effect on the MAD models. Complex structure and sign flips are a generic prediction of these simulations.

although they point north overall (in the positive z -direction). This is a natural consequence of the tangential stretching of vertical field lines as they are dragged into the BH by accreting material, as well as frame dragging (see e.g. Contopoulos et al. 2009; Gabuzda 2018, for helpful schematics). Since this sign flip occurs when crossing the mid-plane, accreting streams of gas that straddle the mid-plane can exhibit streaks of positive RM adjacent to streaks of negative RM.

Returning to Fig. 4 and comparing the 2 MHz bandwidth visualizations to the 4 GHz bandwidth visualizations, SANE models are more strongly affected by bandwidth depolarization than MAD models. In each of the SANE models, counter-jet polarized emission is especially suppressed. This more strongly changes the morphology

of the prograde SANE than the retrograde SANEs, because the image morphologies of the counter-jet and forward-jet in the retrograde SANEs are more similar.

Closely examining the 2 MHz counter-rotating SANE models, one may notice that the linearly polarized intensity appears to vary strongly among adjacent pixels, causing the appearance of ‘static’ across the image. This effect is an artefact of single-frequency radiative transfer calculations, an instance of Faraday rotation randomizing not only the phase of the linear polarization, but also its amplitude. In this model, the linear polarization in each pixel is well approximated by the sum of its forward-jet and counter-jet components, neither of which exhibit this ‘static’ if plotted individually. The

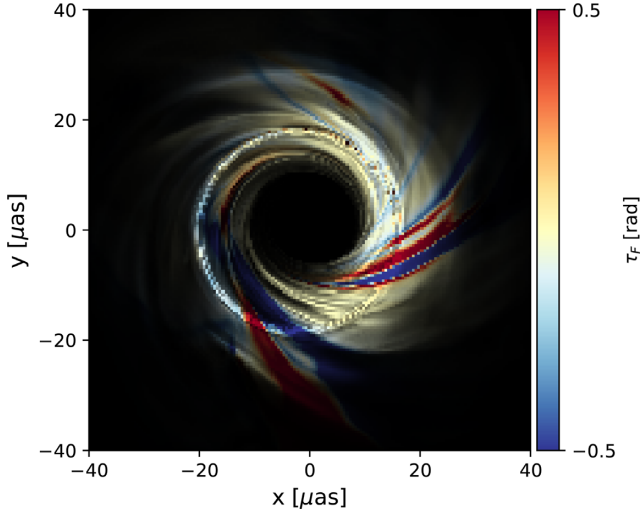


Figure 5. Intensity-weighted Faraday depth in one snapshot of the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model, revealing clear sign flips in the magnetic field parallel to the line of sight. The colour of each pixel encodes the total Faraday depth, while the brightness is proportional to the linear polarized intensity. The sign of the Faraday depth directly encodes the direction of the magnetic field parallel to the geodesic. Subsequent 3D visualization of this snapshot reveals that these sign flips occur in the tangential magnetic field in the plane of the disc.

phase of the counter-jet emission is effectively randomized by the enormous Faraday depth in the mid-plane. Depending on the relative phase, the rotated counter-jet polarized emission may sometimes cancel with the forward-jet polarized emission. This effect would not occur in real observations integrated over a finite bandwidth.

3.2 RM distribution functions for M87*

In these models, polarized emission exhibits a wide range of Faraday depths, and the front and back halves of the emitting regions can differ by many orders of magnitude. In Fig. 6, we plot the distribution functions of $\log_{10}|\text{RM}|$ among the pixels of each model during their final snapshot, weighted by each pixel's linear polarized intensity. This quantity, which we denote as $d|p|/d\log_{10}|\text{RM}|$, is closely related to $F(\phi)$ in RM synthesis theory, the complex polarized surface brightness per unit Faraday depth (Burn 1966; Brentjens & de Bruyn 2005). Unlike $F(\phi)$, we take absolute values, normalize with respect to the total emitted polarized surface brightness, and adopt logarithmically spaced bins.

In each pixel, $|\text{RM}|$ is computed directly from the gradient at 228 GHz across $\Delta\nu = 2$ MHz. These distributions are computed for each of the 11 snapshots that we study, and the filled regions span the range permitted by all of these snapshots. The blue regions only include emission originating from the front half of the emitting region, while the red regions only include emission originating from the back half. The dashed vertical line marks RM_{crit} (eqn. 14) for a 4 GHz bandwidth. Any emission to the right of this line is significantly bandwidth depolarized. The distributions of front half of the emission region can be significantly displaced from that of the back half, especially in SANE models. Again, this is due to the large Faraday depth of the sub-relativistic mid-plane in these models. In some models, including all of the SANEs, the forward-jet distributions exhibit two distinct peaks. The peak at higher $|\text{RM}|$ is due to photon ring orbits, which pass through this mid-plane. Note the extreme difference in $|\text{RM}|$ between front and back components in the

jet-dominated retrograde SANE models. These models may exhibit much lower spatially unresolved $|\text{RM}|$ than would be expected by integrating their Faraday depths across geodesics, since this Faraday depth mainly only affects the counter-jet.

3.3 Non-linear structure of spatially unresolved RM

In Section 3.1, we demonstrate that GRMHD models exhibit rich spatial structure, whereby the RM can vary by many orders of magnitude and flip sign. As a consequence, $\chi(\lambda^2)$ rotates at very different rates at different locations within the image, which may result in clear departures from a λ^2 law if these structures are not spatially resolved. In Fig. 7, we investigate this non-linearity by plotting $\chi(\lambda^2)$ for the final snapshots of our seven models. In blue, we laboriously compute 255 individual images across the 4 GHz bandwidth, then average Stokes parameters within 16 smaller 0.25 GHz bands to estimate $\chi(\lambda^2)$. In red, we instead use the Taylor expansion model based on six images and analytic integrals described in Section 2.2 to estimate $\chi(\lambda^2)$.

All models exhibit significantly non- λ^2 behaviour even within this small fractional bandwidth except for the MAD $R_{\text{high}} = 20$ models. SANE models are especially non-linear, and in fact exhibit spectrally unresolved structure in $\chi(\lambda^2)$ even among the 255 images separated in frequency by 16 MHz. This is because as shown in Fig. 6, a significant amount of the intensity has an associated $|\text{RM}|$ of $\approx 10^9 \text{ rad m}^{-2}$. The MAD $R_{\text{high}} = 160$ models exhibit relatively mild non-linearity, since $|\text{RM}|$ only just approaches $|\text{RM}_{\text{crit}}|$ in Fig. 6.

Recall that we use $\Delta\chi/\Delta\lambda^2$ across the bandwidth from our Taylor expansion model to assign spatially unresolved RMs to images. This figure reveals some of this model's limitations. The model poorly reproduces $\chi(\lambda^2)$ for the MAD, $a = +0.94$, $R_{\text{high}} = 160$ model, possibly due to 228 GHz being a local extremum of $\chi(\lambda^2)$ where the first derivative is small. Interestingly, the retrograde SANE Taylor expansion models appear to broadly follow the structure of the true $\chi(\lambda^2)$, but with a vertical offset. This is due to incorrect evolution of emission superposed on top of the photon ring. Our Taylor expansion model assigns a large $d\phi/d\lambda^2$ to photon ring pixels, but these pixels also contain forward-jet emission that does not pass through the mid-plane. In the SANE, $a = -0.94$, $R_{\text{high}} = 80$ model, this superposed component is immediately bandwidth depolarized and subtracted, leading to the offset. This effect is more delayed as a function of $\Delta\lambda^2$ in the SANE, $a = -0.94$, $R_{\text{high}} = 10$ model, which by construction has a warmer mid-plane and therefore less Faraday rotation. Fortunately, this effect is symmetric about the Taylor expansion point at 228 GHz and we can still recover the spatially unresolved RM from χ at the end points of the band.

3.4 RM as a measure of accretion rate

RM is often used to approximate the accretion rate \dot{M}_\bullet , based on simple analytic models (Marrone et al. 2006). These models are based on advection or convection dominated accretion flows (Narayan & Yi 1994; Narayan, Igumenshchev & Abramowicz 2000; Quataert & Gruzinov 2000a) and make many simplifying assumptions. These include spherical symmetry, equipartition of energy, and rather arbitrary inner and outer radii to truncate the model. Adapted from Marrone et al. (2006),

$$\text{RM} = (3.4 \times 10^{19} \text{ rad m}^{-2}) \left[1 - (r_{\text{out}}/r_{\text{in}})^{-(3\beta-1)/2} \right] \times \left(\frac{\dot{M}_\bullet}{3.5 \times 10^6 M_\odot} \right)^{-2} \left(\frac{2}{3\beta-1} \right) r_{\text{in}}^{-7/4} \dot{M}_\bullet^{3/2}, \quad (16)$$

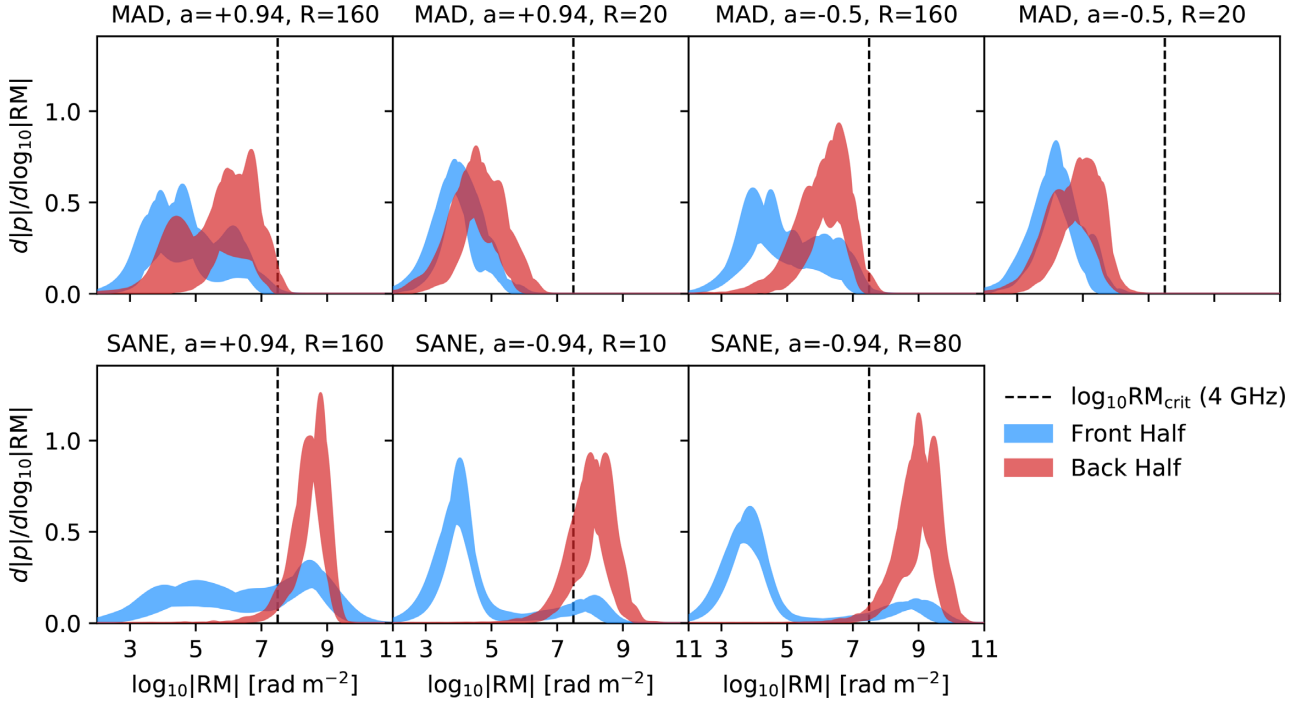


Figure 6. Distribution functions of $\log_{10}|\text{RM}|$ for our models, weighted by the linear polarized intensity of each pixel. Emission originating from front side of the emitting region is coloured blue, while emission originating from the back side is coloured red. The relatively cold mid-plane is the dominant source of Faraday rotation in these models, which can result in significant offsets in these distributions between the two halves, especially in SANE models. Forward-jet emission in SANE models exhibits a second peak at high $\log_{10}|\text{RM}|$ due to photon ring geodesics, which do pass through the mid-plane.

where we have corrected the exponent of r_{in} , as noted by Macquart et al. (2006). Here, r_{out} and r_{in} are the radii used to truncate the model in units of Schwarzschild radii, and $\beta \in [0.5, 1.5]$ describes the slope of the density profile. Note that the model is insensitive to the choice of r_{out} unless $r_{\text{out}}/r_{\text{in}} \approx 1$. \dot{M}_\bullet carries units of $M_\odot \text{ yr}^{-1}$.

Here, we test the relationship between RM and accretion rate in our suite of images for M87*. In Fig. 8, we plot spatially unresolved RM versus accretion rate for the seven models considered in this work. Each symbol of the same colour represents a different snapshot of the same model. Positive RMs are plotted with the filled symbols, while negative RMs are plotted with the open symbols. Notice the ubiquitous flips in the sign of the RM that occur in all models. The blue-filled region demarcates the relation in Marrone et al. (2006), spanning variations of the slope of the density profile, $\beta \in [0.5, 1.5]$. We set $r_{\text{in}} = 3$ and $r_{\text{out}} = \infty$. The dashed line shows the upper limit from Kuo et al. (2014). A 17° inclination appropriate for M87* is shown on the left, while for comparison a 90° inclination is shown on the right.

Overall, we find that a spatially unresolved RM is a poor predictor of the accretion rate, especially if the correct model is not known a priori. RM and the accretion rate differ by orders of magnitude both within and among the different models. Even within a single model, there is no correlation between RM and accretion rate, which we explore in more detail for one model in Section 3.7. Since their higher accretion rates imply higher number densities, SANE models typically have larger spatially unresolved RMs than MAD models. However, these models exhibit such strong time variability that they cannot be distinguished solely by the upper limit of Kuo et al. (2014). Rather, repeat observations on time-scales of months to years will be necessary to characterize the distribution of $|\text{RM}|$ over time and detect potential sign flips.

For a given accretion rate, the GRMHD models in this work produce much lower RM than the analytic model of Marrone et al.

(2006) at an inclination of 17° . As we further explore in Section 3.5, this is because these simulations are viewed through an evacuated funnel region at low inclination. With an inclination of 90° , the $|\text{RM}|$ more closely matches that predicted by Marrone et al. (2006), although they still remain systematically offset. A similar inclination dependence is found for SANE models in Mościbrodzka et al. (2017). The retrograde SANE models remain the most offset from the analytic model. Even at 90° , the large Faraday depth occurs in an area with little emission, since the electrons are assigned low temperatures in the mid-plane.

3.5 Dependence of RM on inclination

Here, we study the dependence of RM on inclination in greater detail. In Fig. 9, we plot the distributions of RM for all 11 snapshots of all seven models at five different inclinations. In this plot, boxes contain the 25th to 75th quantiles, the horizontal black or yellow line marks the median, and the error bars span the full range of the 11 snapshots studied. This plot omits the sign flips observed in Fig. 8, but we comment that they remain ubiquitous at all inclinations for these calculations that terminate at a radius of $20 GM_\bullet/c^2$.

As in Mościbrodzka et al. (2017), we find that the absolute value of the RM depends on the inclination angle, but it is not large compared to the substantial scatter between snapshots. This weaker dependence is likely due to the small radius at which we truncate our calculations. We notice no differences between our calculations at low inclination: 5° , 17° , and 30° . This is fortunate for our study of M87*, as this indicates that we need not be concerned with small deviations from our fiducial inclination of 17° .

At low inclinations, we view the accretion flow through an evacuated funnel region with less Faraday rotating material than at high

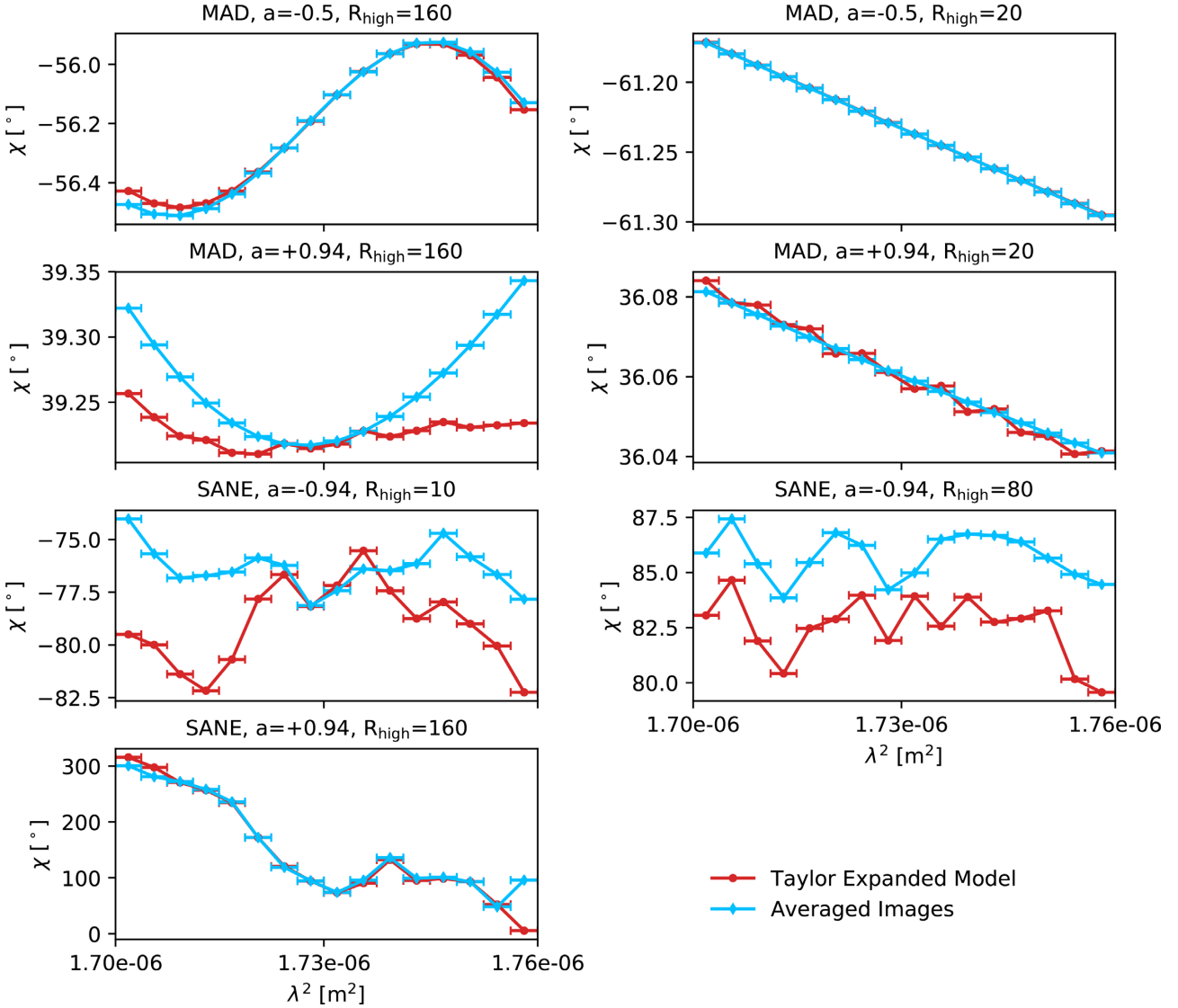


Figure 7. Non- λ^2 behaviour of the seven models we consider during their final simulation snapshot. Within 16 bands that are each 0.25 GHz wide, we plot $\chi(\lambda^2)$ of a band-averaged image. In blue, we plot the result from 255 images evenly spaced between 226 and 230 GHz. Within each band, the appropriate subset of images is averaged. In red, we plot the result from our Taylor expansion model based on six images around 228 GHz. Most of these models exhibit non-linear behaviour even within this narrow fractional bandwidth.

inclinations. We demonstrate this by calculating the characteristic distance of Faraday rotating material as a function of inclination. By modifying the IPOLE source code, we compute for each geodesic

$$\langle R_{\text{FR}} \rangle \equiv \int R_{\text{BL}} |\rho_V| LP ds / \int |\rho_V| LP ds, \quad (17)$$

where ρ_V is the frame-invariant radiative transfer coefficient responsible for Faraday rotation, R_{BL} is the radius of the material in Boyer–Lindquist (or equivalently Kerr–Schild) coordinates, $LP = \sqrt{Q^2 + U^2}$ is the total amount of linearly polarized emission that has been emitted along the geodesic so far (on the way to the camera), and s is the affine parameter describing the geodesic. In other words, this is the characteristic distance of Faraday rotating material, weighted by the fraction of the final linearly polarized emission that has already been added to the pixel on the way to the camera. Once R_{FR} is computed for each pixel, a single value is calculated for the model

by computing an average across the image, weighted by total final linear polarization of each pixel.

In Fig. 10, we plot the characteristic distance of Faraday rotating material of these models during their final snapshot as a function of inclination. For inclinations $< 30^\circ$, most of the Faraday rotation occurs at $< 10 GM_\bullet/c^2$, while $\langle R_{\text{FR}} \rangle$ increases at higher inclinations. The innermost stable circular orbit exists at smaller radius for prograde models than for retrograde models, which leads to a noticeable difference in $\langle R_{\text{FR}} \rangle$ between these two classes of models at low inclination.

Recall that we restrict the domain of our calculations to within $20 GM_\bullet/c^2$, the radius within which the simulations are in inflow equilibrium. As we further explore in Appendix C, if this radius is increased to $50 M_\bullet$, we find consistent results for $i \leq 17^\circ$, but substantially larger $\langle R_{\text{FR}} \rangle$ for $i \geq 60^\circ$. Therefore, we believe that our $|RM|$ values throughout the paper should be considered lower limits for $i > 17^\circ$, as material from beyond the converged region may contribute to Faraday rotation at larger inclinations.

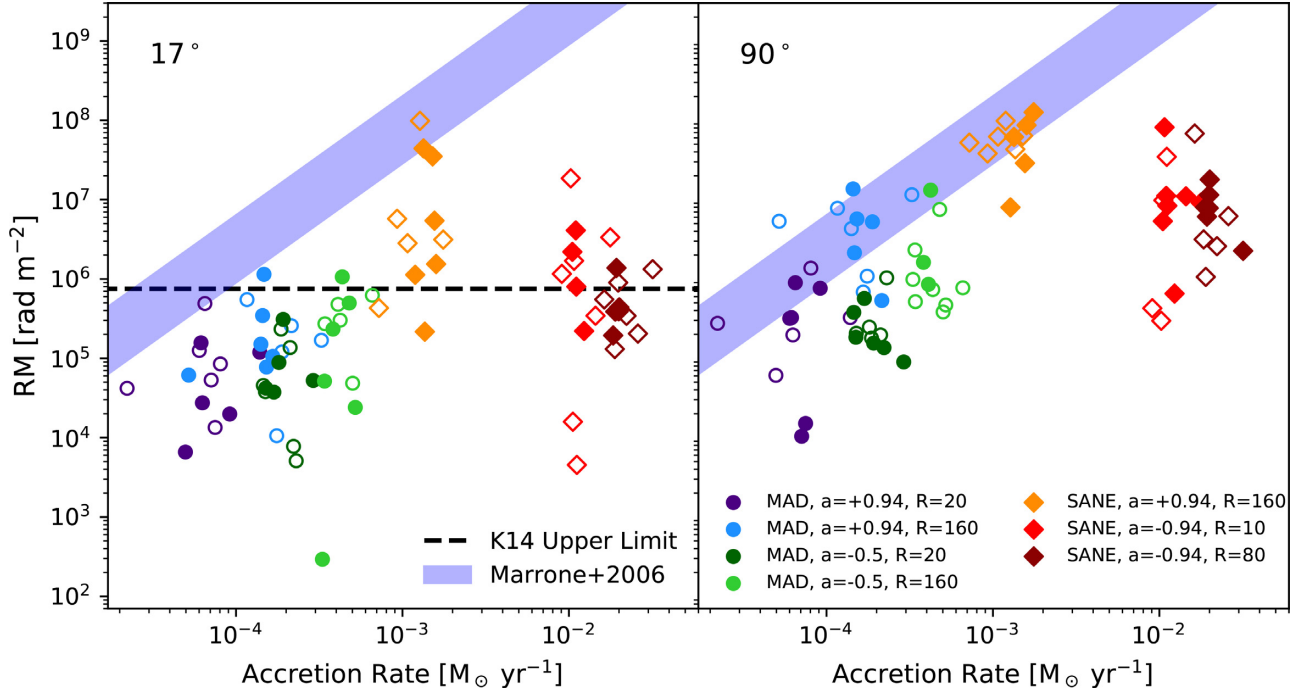


Figure 8. RM as a function of accretion rate for the seven models considered in this paper. The filled symbols have positive RM, while the open symbols have negative RM. In the left-hand panel, the observer is oriented at a 17° inclination, while in the right-hand panel, the observer is oriented at 90° . At an inclination of 17° , we find that accretion rates are systematically higher than those that would be inferred by simple analytic models (Marrone et al. 2006). At 90° , we find RMs in better agreement with analytic models, although there remains substantial scatter. Retrograde SANE models are outliers, since their forward-jet emission does not intercept the large Faraday depth in the mid-plane.

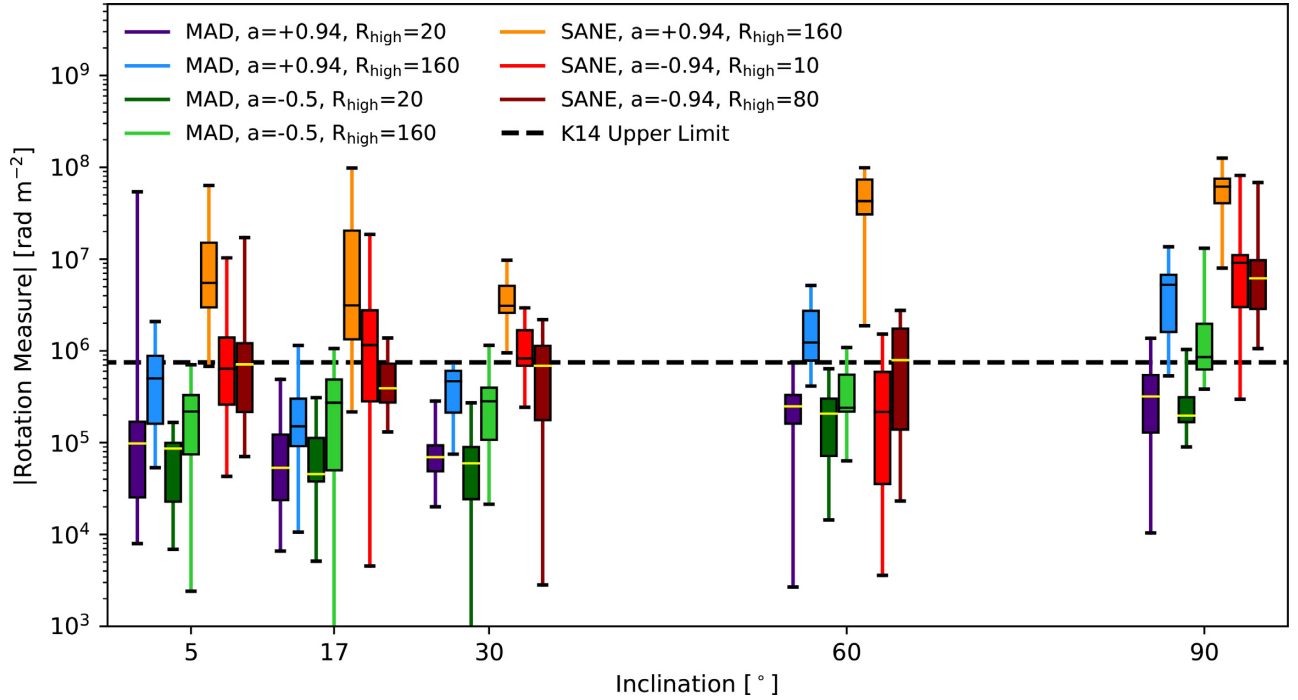


Figure 9. RM as a function of inclination. A total of 11 snapshots are shown for each model. For M87*, 17 deg is considered the most likely inclination based on its large-scale jet. We plot the Kuo et al. (2014) upper limit on the RM with a dashed line. Boxes contain the 25th to 75th quantiles, the horizontal black or yellow lines mark the median, and the error bars span the full range of the 11 snapshots considered. We report a noticeable inclination dependence, due to the evacuated jet region through which the BH is observed at low inclination.

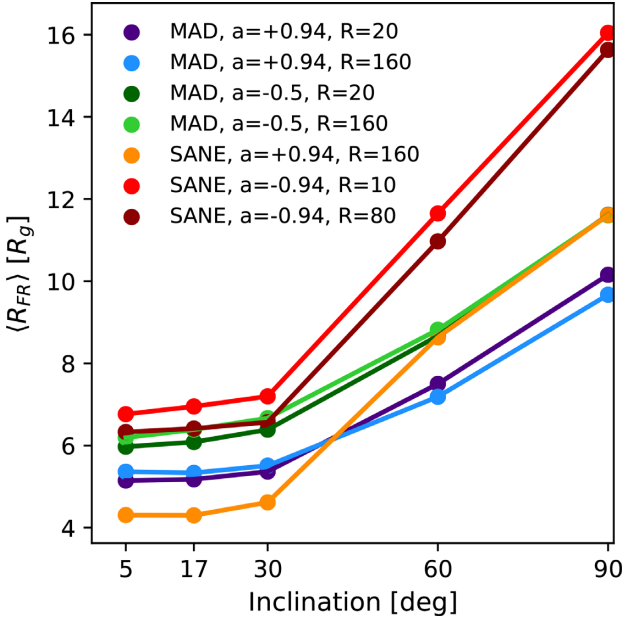


Figure 10. Characteristic distance of Faraday rotating material in these models during their final snapshot as a function of inclination. Due to the evacuated jet region in these simulations, the Faraday rotating material is confined to low radius at low inclination, but extends to larger radius at higher inclination. Recall that our calculations terminate at $R = 20 \text{ GM}_\bullet/c^2$, within which these simulations are in inflow equilibrium.

Since the RM is highly non-uniform across these images, spatially resolved RM distribution functions provide greater insight into the inclination dependence. In Fig. 11, we plot the RM distribution functions as in Fig. 6, for inclinations $i \in \{5^\circ, 30^\circ, 60^\circ, 90^\circ\}$.

Unlike in Fig. 6, we do not split the distributions into their front and back halves. At low inclination, these models exhibit significant emission with low $|\text{RM}|$. As the inclination approaches 90° , this fraction of emission with low $|\text{RM}|$ diminishes, and the distribution is skewed towards higher values. Note that the distributions at 5° are indistinguishable from those at 30° .

3.6 Non-Linear $\chi(\lambda^2)$ as a model discriminant

Since emission and Faraday rotation occur co-spatially and non-uniformly throughout these models, $\chi(\lambda^2)$ need not be linear. We find that the degree of non-linearity varies significantly among models, due to the complex RM structure described in Section 3.1. As a metric of non-linearity, we fit lines to $\chi(\lambda^2)$ from the 16 small bands spanning 226 to 230 GHz and obtain the coefficient of determination, R^2 , defined via

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}. \quad (18)$$

Here, SS_{res} is the regression sum of squares, and SS_{tot} is the total sum of squares. R^2 describes the fraction of variation within the data that can be ascribed to the simple linear dependence.

Our results are shown in Fig. 12, following the same formatting as Fig. 9. Based on these results, SANE models should exhibit non-linear behaviour most of the time. In contrast, MAD models are almost always well described by a linear $\chi(\lambda^2)$ law, especially those with $R_{\text{high}} = 20$. This can be understood by returning to Fig. 6. Pixels with $|\text{RM}| > |\text{RM}|_{\text{crit}}$ have individual EVPAs that rotate substantially across the 4 GHz bandwidth. Images consisting of a substantial fraction of such pixels will therefore exhibit structure in $\chi(\lambda^2)$ within the bandwidth. Among the models considered in this study, this behaviour appears much more likely among SANES.

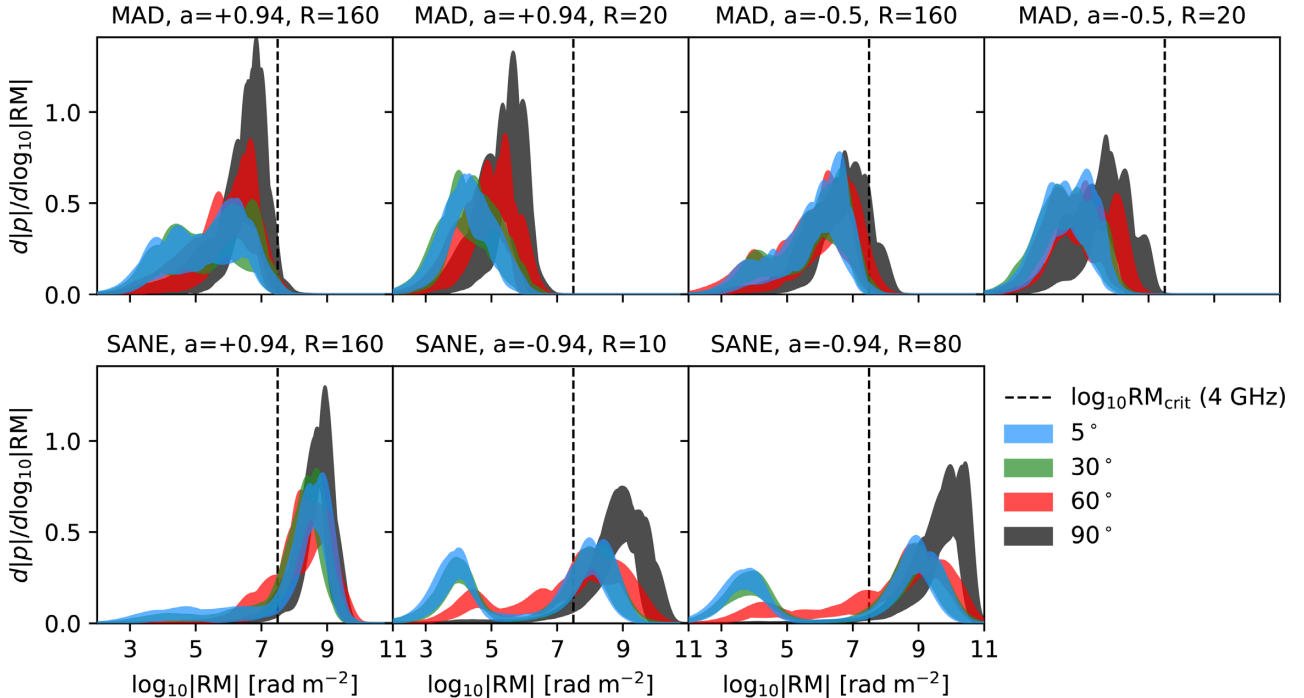


Figure 11. RM distribution functions as in Fig. 6, now shown as a function of inclination. Both halves of the emitting region are combined in this figure. At higher inclinations, we find that these distributions are skewed towards higher values, as the population of photons experiencing comparatively little RM diminishes.

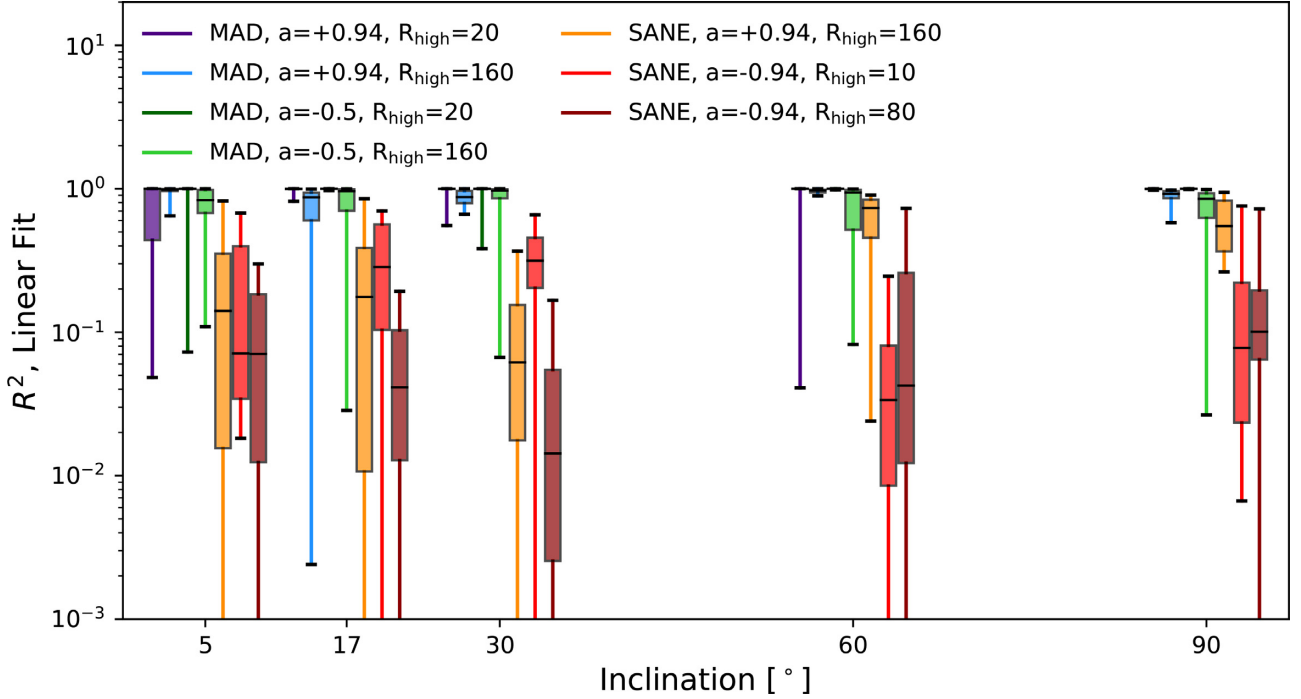


Figure 12. Non-linearity of $\chi(\lambda^2)$ among the models considered in this study. R^2 of a linear fit to $\chi(\lambda^2)$ describes the fraction of variation in the data that can be explained by a simple linear model. SANE models exhibit more non-linearity than MAD models in this study, since their images contain a significant amount of polarized intensity with $|\text{RM}| > |\text{RM}|_{\text{crit}}$.

3.7 Case study: RM time variability

We study one model with higher time resolution in order to quantify the variability of its RM. For the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model, we create images for every available snapshot within $t/(GM_{\bullet}/c^3) \in [7500, 10000]$, which are each separated by $5 GM_{\bullet}/c^3$. This model has the smallest accretion rate and $|\text{RM}|$ of the models explored in this work. Its $|\text{RM}|$ is sufficiently smaller than RM_{crit} that $\chi(\lambda^2)$ remains linear within the bandwidth (see Fig. 7), and thus we create only two polarized images (two frequencies) at each snapshot instead of the usual six (three frequencies, each separately for two sides of the disc).

The time variability of this model at an inclination of 17° is visualized in Fig. 13. In the top row, panels are separated by about half a year, while in the bottom row panels are separated by about 5 d. As in previous figures, pixel brightness encodes the linear polarized intensity, while the colour encodes the RM. Both positive and negative RM regions can be found in a typical snapshot. The spatially unresolved RM is written at the bottom of each panel. The bottom row illustrates how the dynamics of an image with both positive and negative RM regions can result in variability and RM sign flips on a time-scale of a few days. The RM sign flip does not require a drastic change in global source structure; rather, the balance between positive and negative RM regions is shifted.

In Fig. 14, we plot the RM as a function of time for this model, as well as its autocorrelation function. The geometrized time unit is converted to days via $t_{\bullet} = GM_{\bullet}/c^3$, which for M87* is 8.5 hr. The grey band encloses the 16th–84th (1σ) percentiles. At 17° , these include both positive and negative values, such that $\text{RM} = -0.00^{+1.27}_{-0.83} \times 10^5 \text{ rad m}^{-2}$. Examining the images, we do not notice any obvious special behaviour in the accretion flow during periods of large $|\text{RM}|$. Since the RM is determined by the motion of material on event horizon scales, the autocorrelation function of this

time series drops rapidly, falling below 0.5 in less than the separation between snapshots.

In Fig. 15, we plot the joint probability distributions of $\log_{10}|\text{RM}|$ with accretion rate, linear polarized intensity, and circular polarized intensity for the model at 17° . 1σ , 2σ , and 3σ contours are overlaid in white. We do not find any correlation between $|\text{RM}|$ and \dot{M}_{\bullet} , indicating that within a single model, a change in RM does not imply a change in the accretion rate, as might be suggested by analytic models. Rather, as we have discussed, $|\text{RM}|$ and its sign appears to result from a complicated and stochastic cancellation of positive and negative regions. For Sgr A*, Bower et al. (2018) found an anticorrelation between linear polarized intensity and RM, but no such correlation with circular polarized intensity. For this particular model of M87*, we recover qualitatively similar results: a linear regression yields $\log_{10}|\text{RM}| = -1.2\log_{10}LP + 2.7 \pm 0.08$, where $LP = \sqrt{Q^2 + U^2}$ in Jy and RM is in units of rad m^{-2} , with a moderate r-value of -0.57 . No statistically significant correlation is found between $|\text{RM}|$ and circular polarization. An anticorrelation between $|\text{RM}|$ and LP is not surprising, since greater Faraday rotation implies greater scrambling of the polarization vector field.

4 DISCUSSION

4.1 Comparing different models

Here, we summarize the qualitative commonalities and differences between the different models we have considered.

- (i) Prograde MAD: These models require the lowest accretion rate to generate the appropriate total intensity, and consequently exhibit the lowest $|\text{RM}|$. Compared to the other models, there is not too much difference in $|\text{RM}|$ between the two halves of the emitting region,

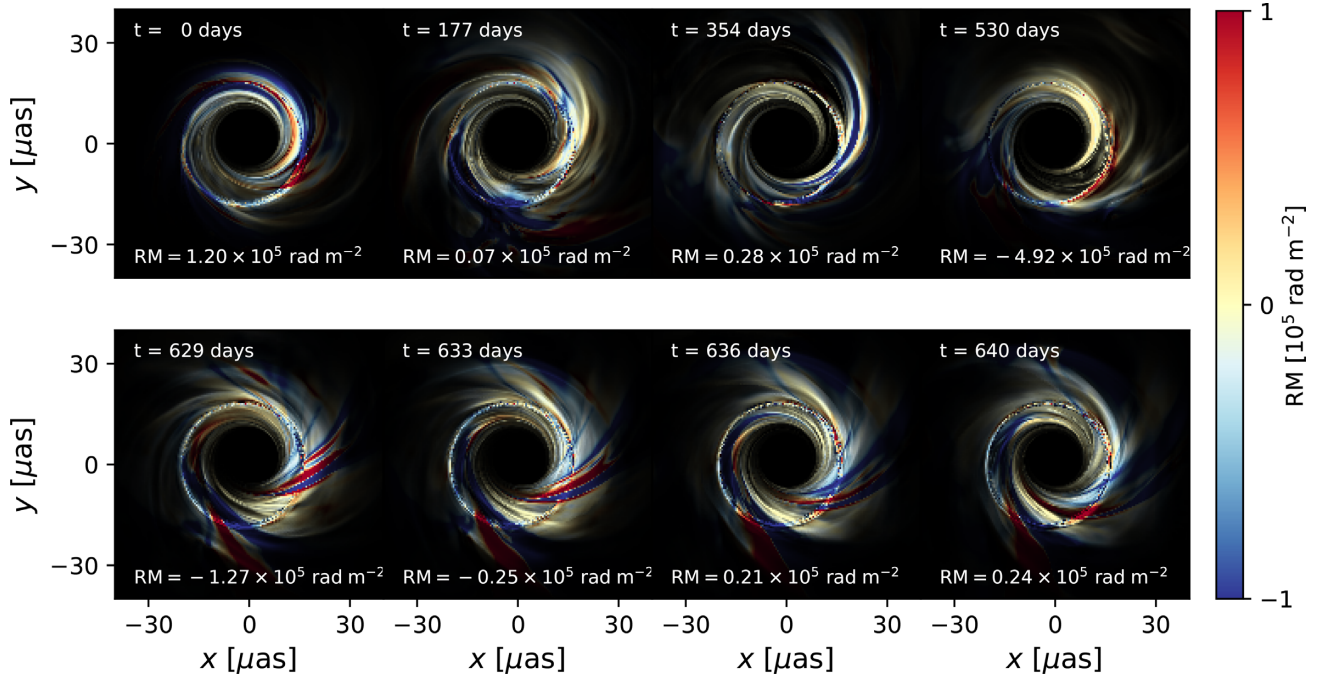


Figure 13. Visualization of the time variability of the RM structure in the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model. The brightness of each pixel scales with its linear polarized intensity, while the colour represents its rotation measure. The RM that would be inferred from a spatially unresolved measurement is written at the bottom of each panel. The spatially unresolved RM can change on time-scales of days as the Faraday rotating gas moves on event horizon scales. The sign flip does not require a dramatic change in the source structure.

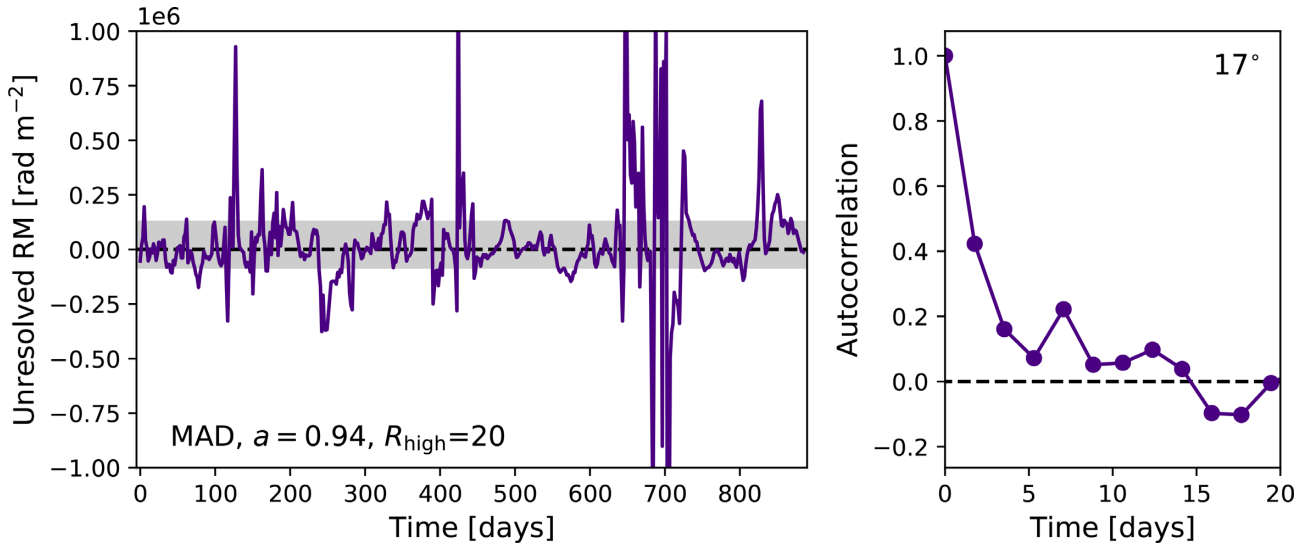


Figure 14. Left: RM as a function of time in the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model that we study with greater time resolution. The grey band encloses the 1σ percentile region over this time, which includes both positive and negative values. Right: Autocorrelation function of this time series. The autocorrelation drops below 50 per cent in less than the separation between snapshots.

since both components occur close to the mid-plane. These models usually exhibit linear $\chi(\lambda^2)$ within 4 GHz.

(ii) Retrograde MAD: Retrograde MAD models require larger accretion rates than their prograde counterparts, but exhibit similar values of $|RM|$. Some areas of the $R_{\text{high}} = 160$ models have large enough $|RM|$ to weakly bandwidth depolarize or scramble portions of the image.

(iii) Prograde SANE: Due to lensing, the counter-jet spans a larger angular scale and contributes more to the total intensity than

the forward-jet. Bandwidth depolarization effects are severe, and emission from the counter-jet is entirely depolarized, assuming a 4 GHz bandwidth. Consequently, the total intensity image (dominated by the counter-jet) appears morphologically different compared to the linearly polarized image (dominated by the forward-jet). $\chi(\lambda^2)$ exhibits strongly non-linear evolution.

(iv) Retrograde SANE: The counter-jet experiences six orders of magnitude more Faraday rotation than the forward-jet. As with their prograde counterparts, $\chi(\lambda^2)$ is highly non-linear. However, the

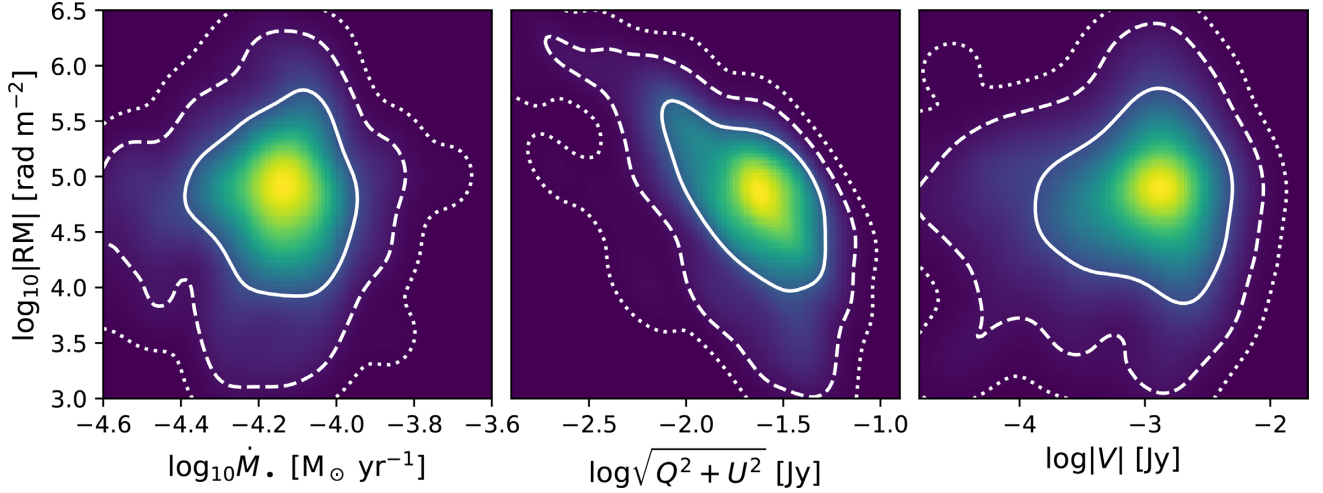


Figure 15. Joint probability distributions of $\log_{10}|RM|$ with accretion rate, linear polarized intensity, and circular polarized intensity in the MAD, $a = 0.94$, $R_{\text{high}} = 20$ model with an inclination of 17° . 1σ , 2σ , and 3σ contours are overlaid in white, using the solid, dashed, and dotted lines, respectively. We find no correlation between $|RM|$ and \dot{M}_\bullet , implying that a change in RM does not imply a change in the accretion rate. As observed by Bower et al. (2018) for Sgr A*, we recover an anticorrelation between $|RM|$ and linear polarization, but not circular polarization.

difference in morphology between the total intensity image and the linearly polarized image is less significant than the prograde case, since the two emission components subtend more similar angular scales.

By construction, increasing R_{high} decreases the temperature of electrons in the mid-plane, which therefore increases the Faraday depth for emission that passes through it. In all models, larger R_{high} results in greater non-linearity of $\chi(\lambda^2)$.

4.2 RM and bandwidth depolarization

If anywhere in an image, $RM > RM_{\text{crit}}$, then that region's linear polarization should be suppressed as described by equation (13). We find that this is more likely to occur in SANE models, but may also affect some parts of MAD models. Bandwidth depolarized regions manifest as areas with lower than average linear polarization fraction.

In some of the worst cases, like that presented in Fig. 2, we find that basic image properties are affected by bandwidth depolarization, such as the linear polarization fraction as well as the morphology of the linear polarization vector field, even after images are blurred. This may be important for studies which use images computed at a single frequency to compare to observations taken over a finite bandwidth. For future studies, the methodology introduced in Section 2.2 can be generalized by applying the Taylor expansion to each point along the ray-traced geodesic instead of just two separate emission regions. This would allow the appropriate bandwidth integrations to occur within the ray-tracing code itself.

4.3 Caveats and limitations

At present, the electron distribution function is poorly constrained. For creating these images, a thermal distribution function is assumed, along with the R_{high} prescription developed by Mościbrodzka et al. (2016). Mao, Dexter & Quataert (2017) studied the effects of adding a power-law component to the distribution function and found that even if a few per cent of the total energy is put into a non-thermal power-law component, a diffuse halo of emission can be produced. The effects of non-thermal electron distribution functions on polarized images remain to be studied.

Recall that we truncate our radiative transfer calculations at a radius of $20 GM_\bullet/c^2$, only within which these GRMHD simulations exhibit inflow equilibrium. However, Faraday rotating material can plausibly exist at larger radius, especially at higher inclinations. Using very long-duration simulations, Dexter et al. (2020) find that Faraday rotation can peak at radii $R \sim 30\text{--}90 GM_\bullet/c^2$, depending on the electron prescriptions. A more distant Faraday screen would be expected to uniformly rotate all EVPAs by a fixed amount, which may leave signatures in the EVPA vector field (Palumbo et al. 2020). Additionally, such a screen should maintain a consistent RM for longer time-scales than these models. Such distant screens may in fact be required to explain the consistent RMs of Sgr A* (Bower et al. 2018) and 3C 84 (Plambeck et al. 2014) over time-scales of years.

5 SUMMARY AND CONCLUSIONS

We have investigated RM within a subset of the EHT simulation library that is consistent with the observational constraints on M87* considered in EHT5. We find more information in the RM structure of GRMHD simulations than can be described by a single scalar, RM. We summarize our results below:

- (i) In a single snapshot, we find extreme variations in the RM between different regions. The RM may vary by orders of magnitude and even flip sign across the image. The RM inferred from a spatially unresolved measurement is therefore the result of the complicated interplay of these different regions.
- (ii) Emission originating from in front of the disc mid-plane may be orders of magnitude less Faraday rotated than emission from the back. In the high accretion rate SANE models, the RM is large enough to completely depolarize emission from the counter-jet, which may in fact dominate the total intensity. The subrelativistic mid-plane is the dominant source of Faraday rotation in these models.
- (iii) Many models exhibit clear departures from a λ^2 law even across a narrow fractional bandwidth of 4 GHz. Non-linearity is a more common feature among the SANE models, and increases with R_{high} .
- (iv) The RM structure changes as material moves on event horizon scales. These models all exhibit strong time variability, causing the

spatially unresolved RM to vary and even flip sign on a time-scale of days.

(v) RM is a poor predictor of the accretion rate. These models predict several orders of magnitude spread in RM for a given accretion rate, and within a single model, these quantities are not correlated as a function of time. In addition, analytic models used to infer the accretion rate based on the RM (Marrone et al. 2006) systematically underestimate the accretion rate on to M87*, since the source should be viewed through an evacuated funnel region.

In future work, a more thorough investigation of the EHT simulation library is merited, including models for Sgr A*. Alternative models for the electron distribution function should also be considered. Repeated observations of both Sgr A* and M87* will be useful to probe the time variability predicted by these models.

ACKNOWLEDGEMENTS

We thank Jason Dexter for valuable feedback that greatly improved the content of this paper. We also thank Laurent Loinard and the anonymous referee for their careful and thorough reviews. This material is based upon work supported by the National Science Foundation under Grant No. OISE 1743747 and NSF AST 1716327. Computations were performed using the resources of the Black Hole Initiative (BHI). Computations at the BHI were made possible through the support of grants from the Gordon and Betty Moore Foundation and the John Templeton Foundation. This work used the Extreme Science and Engineering Discovery Environment resource stampede2 at Texas Advanced Computing Center (TACC) through allocation TG-AST170024. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Moore or Templeton Foundations.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Agol E., 2000, *ApJ*, 538, L121
 Agudo I., Thum C., Ramakrishnan V., Molina S. N., Casadio C., Gómez J. L., 2018, *MNRAS*, 473, 1850
 Bardeen J. M., Wagoner R. V., 1969, *ApJ*, 158, L65
 Bower G. C., Wright M. C. H., Falcke H., Backer D. C., 2003, *ApJ*, 588, 331
 Bower G. C., Dexter J., Markoff S., Rao R., Plambeck R. L., 2017, *ApJ*, 843, L31
 Bower G. C. et al., 2018, *ApJ*, 868, 101
 Brentjens M. A., de Bruyn A. G., 2005, *A&A*, 441, 1217
 Broderick A., Blandford R., 2003, *MNRAS*, 342, 1280
 Broderick A. E., Loeb A., 2006a, *MNRAS*, 367, 905
 Broderick A. E., Loeb A., 2006b, *ApJ*, 636, L109
 Broderick A. E., McKinney J. C., 2010, *ApJ*, 725, 750
 Bromley B. C., Melia F., Liu S., 2001, *ApJ*, 555, L83
 Brunthaler A., Bower G. C., Falcke H., Mellon R. R., 2001, *ApJ*, 560, L123
 Burn B. J., 1966, *MNRAS*, 133, 67
 Chael A., Narayan R., Johnson M. D., 2019, *MNRAS*, 486, 2873
 Chatterjee K. et al., 2020, preprint (arXiv:2002.08386)
 Childs H. et al., 2012, in High Performance Visualization—Enabling Extreme-Scale Scientific Insight. Chapman and Hall/CRC, New York, p. 357
 Contopoulos I., Christodoulou D. M., Kazanas D., Gabuzda D. C., 2009, *ApJ*, 702, L148
 Dexter J., 2016, *MNRAS*, 462, 115
 Dexter J., McKinney J. C., Agol E., 2012, *MNRAS*, 421, 1517

Dexter J. et al., 2020, *MNRAS*, 494, 4168
 Eatough R. P. et al., 2013, *Nature*, 501, 391
 Event Horizon Telescope Collaboration, 2019a, *ApJ*, 875, L1
 Event Horizon Telescope Collaboration, 2019b, *ApJ*, 875, L2
 Event Horizon Telescope Collaboration, 2019c, *ApJ*, 875, L3
 Event Horizon Telescope Collaboration, 2019d, *ApJ*, 875, L4
 Event Horizon Telescope Collaboration, 2019e, *ApJ*, 875, L5
 Event Horizon Telescope Collaboration, 2019f, *ApJ*, 875, L6
 Fragile P. C., Blaes O. M., Anninos P., Salmonson J. D., 2007, *ApJ*, 668, 417
 Frick P., Sokoloff D., Stepanov R., Beck R., 2011, *MNRAS*, 414, 2540
 Gabuzda D., 2018, *Galaxies*, 7, 5
 Gammie C. F., McKinney J. C., Tóth G., 2003, *ApJ*, 589, 444
 Gammie C. F., Shapiro S. L., McKinney J. C., 2004, *ApJ*, 602, 312
 Gardner F. F., Whiteoak J. B., 1966, *ARA&A*, 4, 245
 Gebhardt K., Adams J., Richstone D., Lauer T. R., Faber S. M., Gültekin K., Murphy J., Tremaine S., 2011, *ApJ*, 729, 119
 Himwich E., Johnson M. D., Lupsasca A. r., Strominger A., 2020, *Phys. Rev. D*, 101, 084020
 Hovatta T., O’Sullivan S., Martí-Vidal I., Savolainen T., Tchekhovskoy A., 2019, *A&A*, 623, A111
 Howes G. G., 2010, *MNRAS*, 409, L104
 Igumenshchev I. V., Narayan R., Abramowicz M. A., 2003, *ApJ*, 592, 1042
 Jiménez-Rosales A., Dexter J., 2018, *MNRAS*, 478, 1875
 Johnson M. D. et al., 2015, *Science*, 350, 1242
 Johnson M. D. et al., 2020, *Science Advances*, 6, eaaz1310
 Jones T. W., Odell S. L., 1977, *ApJ*, 214, 522
 Kawazura Y., Barnes M., Schekochihin A. A., 2019, *Proc. Natl. Acad. Sci.*, 116, 771
 Kim J. Y. et al., 2019, *A&A*, 622, A196
 King A. R., Pringle J. E., Hofmann J. A., 2008, *MNRAS*, 385, 1621
 Kuo C. Y. et al., 2014, *ApJ*, 783, L33
 Le Roux E., 1961, *Ann. Astrophys.*, 24, 71
 Li Y.-P., Yuan F., Xie F.-G., 2016, *ApJ*, 830, 78
 Liska M., Hesp C., Tchekhovskoy A., Ingram A., van der Klis M., Markoff S. B., Van Moer M., 2020, *MNRAS*
 Macquart J.-P., Bower G. C., Wright M. C. H., Backer D. C., Falcke H., 2006, *ApJ*, 646, L111
 Mao S. A., Dexter J., Quataert E., 2017, *MNRAS*, 466, 4307
 Marrone D. P., Moran J. M., Zhao J.-H., Rao R., 2006, *ApJ*, 640, 308
 Marrone D. P., Moran J. M., Zhao J.-H., Rao R., 2007, *ApJ*, 654, L57
 Medeiros L., Chan C.-k., Özel F., Psaltis D., Kim J., Marrone D. P., Sadowski A., 2018, *ApJ*, 856, 163
 Melrose D. B., 1997, *J. Plasma Phys.*, 58, 735
 Mościbrodzka M., Gammie C. F., 2018, *MNRAS*, 475, 43
 Mościbrodzka M., Falcke H., Shiokawa H., 2016, *A&A*, 586, A38
 Mościbrodzka M., Dexter J., Davelaar J., Falcke H., 2017, *MNRAS*, 468, 2214
 Narayan R., Yi I., 1994, *ApJ*, 428, L13
 Narayan R., Yi I., 1995a, *ApJ*, 452, 710
 Narayan R., Yi I., 1995b, *ApJ*, 452, 710
 Narayan R., Igumenshchev I. V., Abramowicz M. A., 2000, *ApJ*, 539, 798
 Narayan R., Igumenshchev I. V., Abramowicz M. A., 2003, *PASJ*, 55, L69
 Narayan R., Sądowski A., Penna R. F., Kulkarni A. K., 2012, *MNRAS*, 426, 3241
 Palumbo D. C. M., Wong G. N., Prather B. S., 2020, *ApJ*, 894, 156
 Pasetto A., Carrasco-González C., O’Sullivan S., Basu A., Bruni G., Kraus A., Curiel S., Mack K.-H., 2018, *A&A*, 613, A74
 Plambeck R. L. et al., 2014, *ApJ*, 797, 66
 Porth O. et al., 2019, *ApJS*, 243, 26
 Quataert E., Gruzinov A., 1999, *ApJ*, 520, 248
 Quataert E., Gruzinov A., 2000a, *ApJ*, 539, 809
 Quataert E., Gruzinov A., 2000b, *ApJ*, 545, 842
 Rees M. J., Begelman M. C., Blandford R. D., Phinney E. S., 1982, *Nature*, 295, 17
 Roelofs F., Johnson M. D., Shiokawa H., Doeleman S. S., Falcke H., 2017, *ApJ*, 847, 55
 Ryan B. R., Ressler S. M., Dolence J. C., Gammie C., Quataert E., 2018, *ApJ*, 864, 126

- Rybicki G. B., Lightman A. P., 1986, *Radiative Processes in Astrophysics*. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore
- Sazonov V. N., 1969, *Sov. Astron.*, 13, 396
- Shapiro S. L., Lightman A. P., Eardley D. M., 1976, *ApJ*, 204, 187
- Sharma P., Quataert E., Hammett G. W., Stone J. M., 2007, *ApJ*, 667, 714
- Shcherbakov R. V., 2008, *ApJ*, 688, 695
- Shcherbakov R. V., Penna R. F., McKinney J. C., 2012, *ApJ*, 755, 133
- Sądowski A., Narayan R., Penna R., Zhu Y., 2013, *MNRAS*, 436, 3856
- Sądowski A., Wielgus M., Narayan R., Abarca D., McKinney J. C., Chael A., 2017, *MNRAS*, 466, 705
- Thorne K. S., 1974, *ApJ*, 191, 507
- Tsunetoe Y., Mineshige S., Ohsuga K., Kawashima T., Akiyama K., 2020, *PASJ*, 72, 32
- Volonteri M., Sikora M., Lasota J. P., Merloni A., 2013, *ApJ*, 775, 94
- Walker R. C., Hardee P. E., Davies F. B., Ly C., Junor W., 2018, *ApJ*, 855, 128
- Zavala R. T., Taylor G. B., 2004, *ApJ*, 612, 749

APPENDIX A: SPHERICAL STOKES PARAMETERS

In Section 2.2, we convert the standard Stokes parameters $\{Q, U, V\}$ to the spherical Stokes parameters $\{N, \phi, \psi\}$ as in Shcherbakov et al. (2012) for the purposes of a more stable Taylor expansion. This transformation is defined by

$$Q = N \cos \phi \sin \psi, \quad (\text{A1})$$

$$U = N \sin \phi \sin \psi, \quad (\text{A2})$$

$$V = N \cos \psi, \quad (\text{A3})$$

while its inversion can be derived from trigonometric identities as

$$N = \sqrt{Q^2 + U^2 + V^2}, \quad (\text{A4})$$

$$\phi = \arctan(U/Q), \quad (\text{A5})$$

$$\psi = \arctan(\sqrt{Q^2 + U^2}/V). \quad (\text{A6})$$

From these equations, we can see that N is the total amount of both linear and circular polarization, ϕ describes the phase of the linear polarization, and ψ describes the linear to circular polarization ratio. Note the lack of a factor of 1/2 in ϕ , such that $\phi = 2\chi$, where χ is the EVPA.

In a pixel with a large $|\text{RM}|$, Faraday rotation causes Q and U to oscillate rapidly with frequency. In spherical Stokes parameters, N remains stable, while ϕ changes linearly with wavelength squared, with $d\phi/d\lambda^2 = 2\text{RM}$. This makes the Spherical Stokes parameters more stable to Taylor expansion.

When we calculate derivatives of the $\{N, \phi, \psi\}$, we compute them in terms of $\{Q, U, V\}$ and their derivatives. This allows us to avoid mistakes due to phase wrapping, as discussed in Section 2.2. By simply differentiating according to the chain rule, the derivatives are given by

$$N' = \frac{QQ' + UU' + VV'}{\sqrt{Q^2 + U^2 + V^2}}, \quad (\text{A7})$$

$$\phi' = \frac{U'Q - Q'U}{Q^2 + U^2}, \quad (\text{A8})$$

$$\psi' = \frac{V(QQ' + UU') - V'(Q^2 + U^2)}{(Q^2 + U^2 + V^2)\sqrt{Q^2 + U^2}}, \quad (\text{A9})$$

where $'$ denotes differentiation with respect to frequency (or another physically similar quantity such as wavelength or wavelength-squared).

APPENDIX B: ANALYTIC BANDWIDTH INTEGRALS

Here, we provide analytic solutions to the integrals described in Section 2.2. Spherical Stokes parameters $\{I_0, N_0, \phi_0, \psi_0\}$ and their derivatives $\{dI/d\nu, dN/d\nu, d\phi/d\lambda^2, d\psi/d\lambda^3\}$ are estimated at frequency ν_0 . Let ν_c be the central frequency of a band extending between $\nu_1 = \nu_c - \Delta\nu/2$ and $\nu_2 = \nu_c + \Delta\nu/2$. We then define a dimensionless frequency $x = (\nu - \nu_0)/\Delta\nu$ such that $x_1 = x(\nu_1)$ and $x_2 = x(\nu_2)$. We can then write

$$I_{\text{BW}} = \int_{x_1}^{x_2} (I_0 + I_1 x) dx, \quad (\text{B1})$$

$$Q_{\text{BW}} = \int_{x_1}^{x_2} (N_0 + N_1 x) \cos(\phi_0 + \phi_1 x) \sin(\psi_0 + \psi_1 x) dx, \quad (\text{B2})$$

$$U_{\text{BW}} = \int_{x_1}^{x_2} (N_0 + N_1 x) \sin(\phi_0 + \phi_1 x) \sin(\psi_0 + \psi_1 x) dx, \quad (\text{B3})$$

$$V_{\text{BW}} = \int_{x_1}^{x_2} (N_0 + N_1 x) \cos(\psi_0 + \psi_1 x) dx, \quad (\text{B4})$$

where $x_1 = (\nu_c - \nu_0 - \Delta\nu/2)/\Delta\nu$, $x_2 = (\nu_c - \nu_0 + \Delta\nu/2)/\Delta\nu$, and we define the following quantities from the derivatives of the spherical Stokes parameters:

$$I_1 = \frac{dI}{d\nu} \Delta\nu, \quad (\text{B5})$$

$$N_1 = \frac{dN}{d\nu} \Delta\nu, \quad (\text{B6})$$

$$\phi_1 = \frac{-2c^2 \Delta\nu}{\nu_0^3} \frac{d\phi}{d\lambda^2}, \quad (\text{B7})$$

$$\psi_1 = \frac{-3c^3 \Delta\nu}{\nu_0^4} \frac{d\psi}{d\lambda^3}. \quad (\text{B8})$$

These integrals have analytic solutions given by

$$I_{\text{BW}} = I_0(x_2 - x_1) + \frac{I_1}{2} (x_2^2 - x_1^2), \quad (\text{B9})$$

$$\begin{aligned} Q_{\text{BW}} = \frac{1}{2} \Big[& N_0 \left(\frac{\cos(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{\phi_1 - \psi_1} \right. \\ & - \frac{\cos(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{\phi_1 - \psi_1} \\ & + \frac{\cos(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{\phi_1 + \psi_1} \\ & - \frac{\cos(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{\phi_1 + \psi_1} \Big) \\ & + N_1 \left(\frac{\sin(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{(\phi_1 - \psi_1)^2} \right. \\ & - \frac{x_1(\phi_1 - \psi_1) \cos(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{(\phi_1 - \psi_1)^2} \\ & + \frac{x_1(\phi_1 + \psi_1) \cos(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\ & - \frac{\sin(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\ & \left. + \frac{x_2(\phi_1 - \psi_1) \cos(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{(\phi_1 - \psi_1)^2} \right) \end{aligned}$$

$$\begin{aligned}
& - \frac{\sin(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{(\phi_1 - \psi_1)^2} \\
& + \frac{\sin(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\
& - \frac{x_2(\phi_1 + \psi_1) \cos(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \Bigg], \quad (B10)
\end{aligned}$$

$$\begin{aligned}
U_{\text{BW}} = & \frac{1}{2} N_0 \Bigg[- \frac{\sin(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{\phi_1 - \psi_1} \\
& + \frac{\sin(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{\phi_1 + \psi_1} \\
& + \frac{\sin(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{\phi_1 - \psi_1} \\
& - \frac{\sin(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{\phi_1 + \psi_1} \Bigg] \\
& + \frac{1}{2} N_1 \Bigg[- \frac{x_1(\phi_1 - \psi_1) \sin(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{(\phi_1 - \psi_1)^2} \\
& + \frac{\cos(\phi_0 + \phi_1 x_1 - \psi_0 - \psi_1 x_1)}{(\phi_1 - \psi_1)^2} \\
& + \frac{x_1(\phi_1 + \psi_1) \sin(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\
& + \frac{\cos(\phi_0 + x_1(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\
& + \frac{x_2(\phi_1 - \psi_1) \sin(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{(\phi_1 - \psi_1)^2} \\
& + \frac{\cos(\phi_0 + \phi_1 x_2 - \psi_0 - \psi_1 x_2)}{(\phi_1 - \psi_1)^2} \\
& - \frac{x_2(\phi_1 + \psi_1) \sin(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \\
& + \frac{\cos(\phi_0 + x_2(\phi_1 + \psi_1) + \psi_0)}{(\phi_1 + \psi_1)^2} \Bigg], \quad (B11)
\end{aligned}$$

$$\begin{aligned}
V_{\text{BW}} = & \frac{1}{\psi_1^2} \Bigg[- \psi_1(N_0 + N_1 x_1) \sin(\psi_0 + \psi_1 x_1) \\
& + \psi_1(N_0 + N_1 x_2) \sin(\psi_0 + \psi_1 x_2) \\
& - N_1 \cos(\psi_0 + \psi_1 x_1) + N_1 \cos(\psi_0 + \psi_1 x_2) \Bigg]. \quad (B12)
\end{aligned}$$

APPENDIX C: EFFECTS OF CHANGING THE MAXIMUM INTEGRATION RADIUS

In IPOLE, the parameter `rmax_geo` (henceforth R_{out}) sets the radius in Boyer–Lindquist (or Kerr–Schild) coordinates within which radiative transfer coefficients are calculated. Although the MAD and SANE simulations have outer boundaries of $10^3 GM_\bullet/c^2$ and $50 GM_\bullet/c^2$, respectively, we find that these simulations exhibit inflow equilibrium only within $20 GM_\bullet/c^2$. Hence, for the results throughout this paper, R_{out} is therefore set to $20 GM_\bullet/c^2$. Here, we explore the effects of changing this outer radius to $50 GM_\bullet/c^2$. Although this choice now includes material from unconverged regions, this allows us to gain some insight into how gas in more distant regions might affect our predictions.

In Fig. C1, we compare the characteristic distance of Faraday rotating material, $\langle R_{\text{FR}} \rangle$, as in Fig. 10. The bold solid lines originate from the $R_{\text{out}} = 20 GM_\bullet/c^2$ models (as shown in Fig. 10), while the faint thin lines originate from the $R_{\text{out}} = 50 GM_\bullet/c^2$ models. Interestingly, there is little difference in $\langle R_{\text{FR}} \rangle$ for inclinations $i \leq 17^\circ$, an evacuated funnel region in these simulations. At larger inclinations, distant material from unconverged regions could potentially contribute the majority of the Faraday rotation.

In Fig. C2, we plot the RM distribution functions as in Fig. 11, where our $R_{\text{out}} = 20 GM_\bullet/c^2$ results are shown as the solid lines and alternative $R_{\text{out}} = 50 GM_\bullet/c^2$ results are plotted as the dotted lines. For clarity, we only plot the median values at a given $\log_{10} |\text{RM}|$ instead of the full range plotted in Fig. 11. We find that there is negligible difference in our results for inclinations $i \leq 30^\circ$, the range relevant for M87*. At higher inclinations, the MAD distributions are skewed towards higher values for $i \geq 60^\circ$, while the retrograde SANE models only differ at $i = 90^\circ$. The prograde SANE model shows negligible difference between $R_{\text{out}} = 20 GM_\bullet/c^2$ and $R_{\text{out}} = 50 GM_\bullet/c^2$ even at $i = 90^\circ$.

In Fig. C3, we plot the effect this has on the spatially unresolved RM observed for these sources. Here, the solid boxes correspond to the $R_{\text{out}} = 20 GM_\bullet/c^2$ models (as in Fig. 9), while the faint boxes correspond to $R_{\text{out}} = 50 GM_\bullet/c^2$ models. Lines demarcating medians have been removed for clarity. As expected, there is no noticeable difference for inclinations $i \leq 30^\circ$. For inclinations of $\geq 60^\circ$, the RMs of $R_{\text{out}} = 50 GM_\bullet/c^2$ models can be a factor of a few to orders of magnitude larger, depending on the model.

Finally, we notice that at inclinations of 60° , RM sign flips occur less frequently for the $R_{\text{out}} = 50 GM_\bullet/c^2$ models than for the $R_{\text{out}} = 20 GM_\bullet/c^2$ models. This is to be expected, since material at larger radii evolves on longer time-scales.

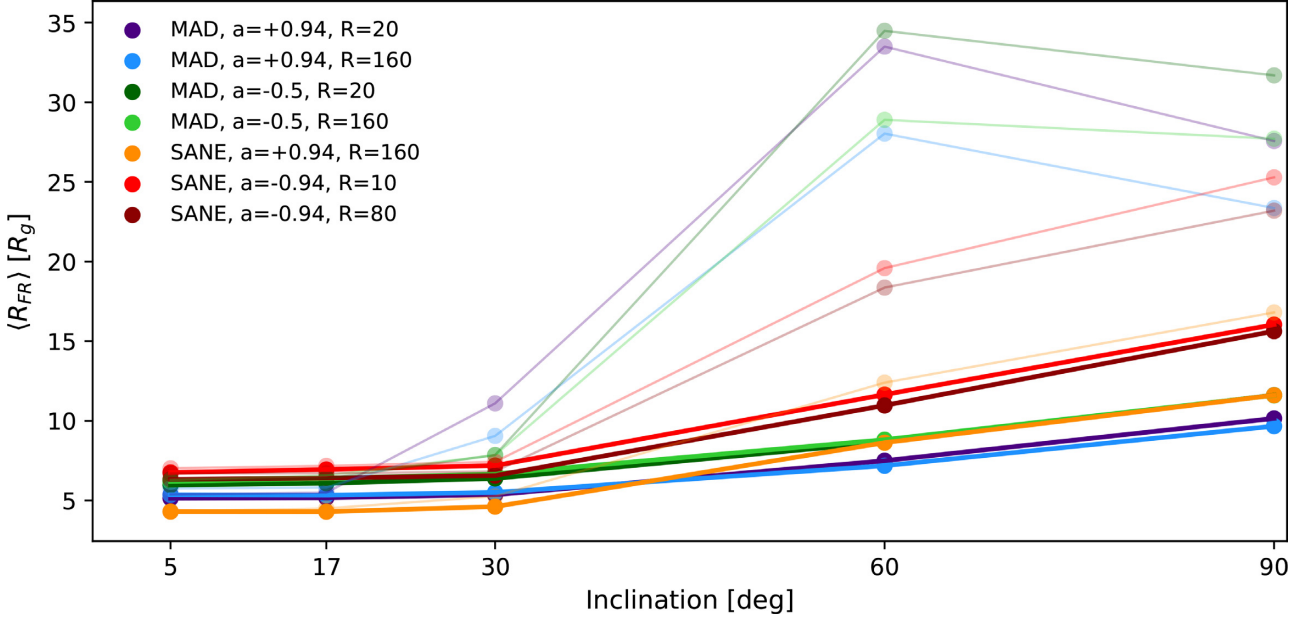


Figure C1. Characteristic distance of Faraday rotating material, as in Fig. 10, for $R_{out} = 20 GM_\bullet/c^2$ models as the solid lines, and $R_{out} = 50 GM_\bullet/c^2$ models as the faint lines. There is little difference for inclinations $i \leq 17^\circ$, where the BH is viewed through an evacuated funnel region. In contrast, material from more distant, unconverged regions can dominate the Faraday rotation at higher inclinations if calculations are allowed to proceed into this area.

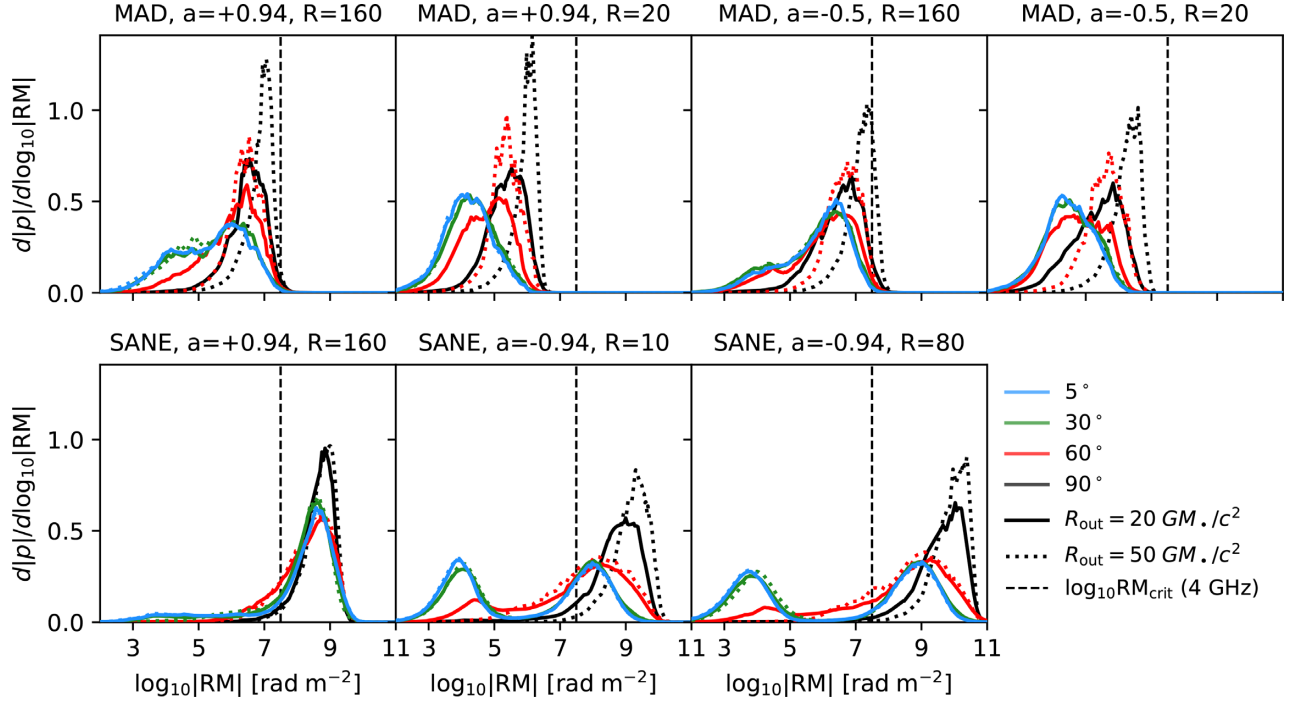


Figure C2. Rotation measure distribution functions, as in Fig. 6, where $R_{out} = 20 GM_\bullet/c^2$ models are shown as the solid lines and $R_{out} = 50 GM_\bullet/c^2$ models are shown as the dotted lines. For clarity, only median values at a given $\log_{10}|RM|$ are plotted. There is negligible difference for all models when the inclination $\leq 30^\circ$. The distributions are skewed towards higher values at 90° for the retrograde SANEs, and for $\geq 60^\circ$ for MADs.

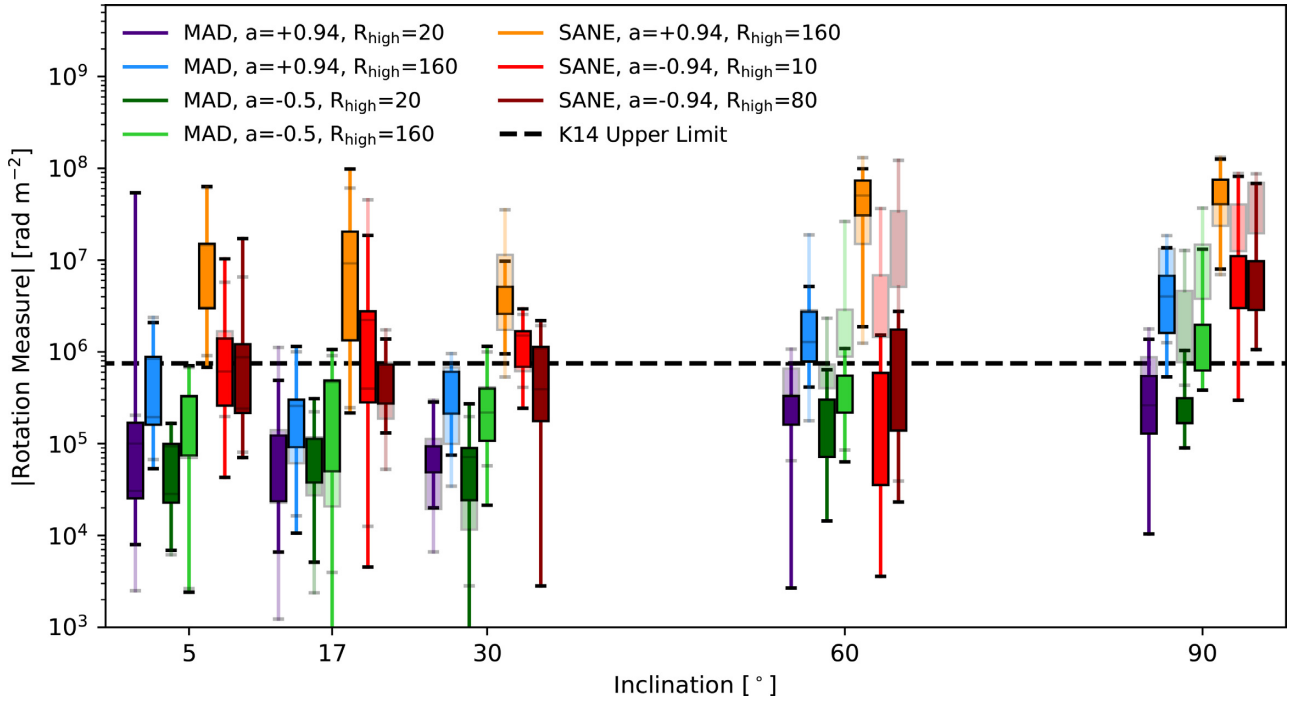


Figure C3. Rotation measure as a function of inclination, as in Fig. 9, for $R_{\text{out}} = 20 \text{ GM}_\bullet/c^2$ models as the solid boxes, and $R_{\text{out}} = 50 \text{ GM}_\bullet/c^2$ models as the faint boxes. While including material at $R_{\text{out}} > 20 \text{ GM}_\bullet/c^2$ makes little difference for inclinations $i \leq 30^\circ$, it may increase the RM by factors of a few to orders of magnitude at larger inclinations, depending on the model.

This paper has been typeset from a \LaTeX file prepared by the author.