The Astrophysical Journal Letters, 930:L16 (49pp), 2022 May 10 https://doi.org/10.3847/2041-8213/ac6672

First Sagittarius A* Event Horizon Telescope Results. V. Testing Astrophysical Models of the Galactic Center Black Hole

The Event Horizon Telescope Collaboration
(See the end matter for the full list of authors.)
Received 2022 March 7; revised 2022 April 7; accepted 2022 April 7; published 2022 May 12

Abstract

In this paper we provide a first physical interpretation for the Event Horizon Telescope's (EHT) 2017 observations of Sgr A*. Our main approach is to compare resolved EHT data at 230 GHz and unresolved non-EHT observations from radio to X-ray wavelengths to predictions from a library of models based on time-dependent general relativistic magnetohydrodynamics simulations, including aligned, tilted, and stellar-wind-fed simulations; radiative transfer is performed assuming both thermal and nonthermal electron distribution functions. We test the models against 11 constraints drawn from EHT 230 GHz data and observations at 86 GHz, 2.2 μm, and in the X-ray. All models fail at least one constraint. Light-curve variability provides a particularly severe constraint, failing nearly all strongly magnetized (magnetically arrested disk (MAD)) models and a large fraction of weakly magnetized models. A number of models fail only the variability constraints. We identify a promising cluster of these models, which are MAD and have inclination $i \leqslant 30°$. They have accretion rate $(5.2–9.5) \times 10^{-9}\ M_{\odot}\ \mathrm{yr}^{-1}$, bolometric luminosity $(6.8–9.2) \times 10^{35}\ \mathrm{erg\ s}^{-1}$, and outflow power $(1.3–4.8) \times 10^{38}\ \mathrm{erg\ s}^{-1}$. We also find that all models with $i \geqslant 70°$ fail at least two constraints, as do all models with equal ion and electron temperature; exploratory, nonthermal model sets tend to have higher 2.2 μm flux density; and the population of cold electrons is limited by X-ray constraints due to the risk of bremsstrahlung overproduction. Finally, we discuss physical and numerical limitations of the models, highlighting the possible importance of kinetic effects and duration of the simulations.

Unified Astronomy Thesaurus concepts: Black hole physics (159); Galactic center (565)

1. Introduction

The center of the Milky Way contains a massive compact object that is likely a supermassive black hole (Do et al. 2019; Gravity Collaboration et al. 2019). The putative black hole is surrounded by hot plasma that is visible across 17 decades in electromagnetic frequency. Hereafter we will use Sgr A* to refer to the supermassive black hole candidate and the hot plasma.

Sgr A* is one of the most studied objects on the sky, both observationally and theoretically. A key characteristic of the Sgr A* system is its extremely low overall luminosity with respect to the Eddington limit. The low luminosity suggests that matter falls onto Sgr A*'s central object in the form of a radiatively inefficient/advection-dominated accretion flow (RIAF/ADAF, as proposed by Ichimaru 1977; Rees et al. 1982; Narayan & Yi 1994, 1995a, 1995b; Narayan et al. 1996, 1998; Yuan & Narayan 2014) rather than in the form of a radiatively efficient thin disk (Shakura & Sunyaev 1973). Since the nearly flat radio spectrum of Sgr A* is similar to radio spectra observed in jets from active galactic nuclei, it has also been suggested that the majority of the Sgr A* emission could be produced by a jet launched by an accreting black hole rather than matter falling through the black hole event horizon (Falcke et al. 1993; Falcke & Markoff 2000).

Models of magnetized RIAFs/ADAFs have been constructed using semianalytic prescriptions (e.g., Narayan et al. 1995; Özel et al. 2000; Broderick et al. 2009, 2011) and using time-dependent general relativistic magnetohydrodynamics (GRMHD) simulations (e.g., Hawley 2000; De Villiers & Hawley 2003; Gammie et al. 2003; Giacomazzo & Rezzolla 2007; Fragile et al. 2012, 2014; White et al. 2016; Anninos et al. 2017; Olivares Sánchez et al. 2018; Olivares et al. 2019; Porth et al. 2019; Liska et al. 2019). Semianalytic RIAF/ADAF models typically do not include relativistic jets or outflows, but those are naturally produced in GRMHD simulation and contribute to the observed emission. GRMHD simulations also naturally produce variability, which is observed in Sgr A* at multiple wavelengths.

GRMHD simulations of ADAFs show that ADAF-like inflows are not unique. In particular, two dramatically different modes are observed, depending on the magnetic flux interior to the black hole equator: the standard and normal evolution (SANE) mode, in which the midplane magnetic field pressure is less than the gas pressure and magnetic fields are turbulent, and the magnetically arrested disk (MAD) mode, in which magnetic fields are strong and organized and can even disrupt accretion. An outstanding question about Sgr A* is whether the flow is in MAD or SANE mode, or possibly in a third mode that results from wind-fed accretion (Ressler et al. 2020b).

The energy distribution of electrons in the emitting plasma is also not known. Because emission is driven by the synchrotron process, this is critical in determining the observational appearance of the source. In particular, the energy per electron may increase with latitude in the flow, leading to a jet or outflow that outshines an equatorial inflow.

The question of whether emission is dominated by an inflow or outflow is intimately tied to the problem of what drives an outflow, if there is one. In GRMHD simulations of black hole accretion the strength of the outflow depends sensitively on the black hole spin (e.g., Event Horizon Telescope Collaboration et al. 2019c, hereafter M87* Paper V; Narayan et al. 2022). At

large spin GRMHD simulations produce powerful jets driven by extraction of black hole spin energy via the Blandford & Znajek (1977) process. A spatially resolved study of Sgr A* may thus also constrain the black hole spin and provide direct evidence for black hole energy extraction.

Previously published GRMHD models of Sgr A* generically predict source sizes at millimeter wavelengths consistent with observational data (e.g., Doeleman et al. 2008; Mościbrodzka et al. 2009; Dexter et al. 2009, 2010); the radio spectral shape is similar to jet emission (e.g., Mościbrodzka & Falcke 2013; Ressler et al. 2017), and the source linear polarization requires strongly magnetized flow or nonthermal electrons (Johnson et al. 2015; Gold et al. 2017; Dexter et al. 2020).

A major difficulty in determining the nature of Sgr A* radio emission is caused by the interstellar scattering screen that distorts our view of the Galactic center up to $\lambda \sim 1$ mm wavelengths (see Johnson et al. 2018; Psaltis et al. 2018; Issaoun et al. 2019, and references therein). The Event Horizon Telescope (EHT) is a very long baseline interferometric (VLBI) experiment operating at 230 GHz or wavelength $\lambda = 1.3$ mm (see Event Horizon Telescope Collaboration et al. 2019b, hereafter M87* Paper II, for an introduction to the instrument). EHT operates at high enough frequency to penetrate the scattering screen, with angular resolution sufficient to directly image structures in the immediate vicinity of the black hole event horizon.

In 2017 April the EHT observed Sgr A* (among other sources, including the core of the M87 galaxy; see Event Horizon Telescope Collaboration et al. 2019a, hereafter M87* Paper I) and produced the first ever horizon-scale images of the source. We report the results of these observations in Event Horizon Telescope Collaboration et al. (2022a, hereafter Paper II) and Event Horizon Telescope Collaboration et al. (2022b, hereafter Paper III), characterize the basic properties of the emission visible in the EHT images in Event Horizon Telescope Collaboration et al. (2022c, hereafter Paper IV), and discuss implications for tests of general relativity in Event Horizon Telescope Collaboration et al. (2022d, hereafter Paper VI). The main goal of this paper (Paper V in the series) is to provide the first comprehensive physical interpretation of the EHT 2017 Sgr A* data sets.

This paper is structured as follows. Section 2 describes our main assumptions, a one-zone source model, and a standard simulation and synthetic image library used to model near-horizon emission from Sgr A*. Our model library assumes that general relativity is valid and the spacetime around Sgr A* is described by the Kerr metric (Kerr 1963). A discussion of Sgr A* observations in the context of alternative theories of gravity can be found in Paper VI. Our model library is based on time-dependent GRMHD simulations that, combined with general relativistic radiative transfer models, result in images and broadband spectra of the models. The library of simulated images was used in Paper III and Paper IV, to validate the Sgr A* EHT imaging and parameter estimation algorithms. In Section 3, we describe the observational constraints that are used in the present work to test theoretical models of Sgr A*. These data compose a subset of EHT 2017 observations and other non-EHT historical or other data. In Section 4, we describe model scoring procedures and use our model library to infer physical properties of the Sgr A* system. We discuss model limitations, results in the context of previous studies, and the outlook for future Sgr A* theoretical research directions in Section 5. Finally, we conclude in Section 6.

This paper is supplemented with several appendices. Appendix A discusses numerical details of our simulations. Appendix B discusses the impact of physical and numerical effects on the model variability. Appendix C summarizes the results of applying constraints to our fiducial models in an extended set of figures.

2. Astrophysical Models

2.1. Basic Assumptions

We assume that the mass of and distance to Sgr A* are

$$M = 4.14 \times 10^6\, M_\odot, \tag{1}$$

$$D = 8.127\ \mathrm{kpc}, \tag{2}$$

which are approximately the mean of the values reported by Do et al. (2019) and Gravity Collaboration et al. (2019), which differ from each other by about 4%. The distance is consistent with maser parallax measurements (Reid et al. 2019).

We also assume that Sgr A* is a black hole described by the Kerr metric. The dimensionless spin, $a_* \equiv Jc/GM^2$, is a free parameter with $-1 < a_* < 1$, where J, G, and c are the black hole spin angular momentum, gravitational constant, and speed of light, respectively. Following M87* Paper V, we use $a_* > 0$ to indicate that the angular momenta of the accretion flow and black hole are parallel (the accretion flow is "prograde") and $a_* < 0$ to indicate that the angular momenta of the accretion flow and black hole are antiparallel ("retrograde").[152]

The implied characteristic length

$$r_g \equiv GM/c^2 \simeq 6.1 \times 10^{11}\,\mathrm{cm}, \tag{3}$$

the characteristic time

$$t_g \equiv GM/c^3 \simeq 20.4\ \mathrm{s}, \tag{4}$$

and the angular scale

$$\vartheta_g \equiv GM/(c^2 D) \simeq 5.03\ \mu\mathrm{as}. \tag{5}$$

The expected diameter of the black hole shadow is $2\sqrt{27}\,GM/(c^2D)$ for $a_* = 0$. For $|a_*| > 0$ the shadow is noncircular and its size and shape depend on a_* and inclination i (the angle between the line of sight and the spin axis); its width can be as small as $9GM/(c^2D)$ for $a_* = 1$ and $i = 90°$ (Bardeen 1973).

If the emitting plasma is ionized hydrogen, then the Eddington luminosity

$$L_{\mathrm{Edd}} \equiv 4\pi GMcm_p/\sigma_{\mathrm{T}} = 5.2 \times 10^{44}\ \mathrm{erg\ s^{-1}}, \tag{6}$$

where symbols have their usual meaning. The corresponding Eddington accretion rate

$$\begin{aligned}\dot{M}_{\mathrm{Edd}} \equiv L_{\mathrm{Edd}}/(0.1c^2) &= 5.8 \times 10^{24}\ \mathrm{g\ s^{-1}} \\ &= 0.09\, M_\odot\ \mathrm{yr^{-1}},\end{aligned} \tag{7}$$

where the nominal efficiency is 10%. The bolometric luminosity of Sgr A* is $L_{\mathrm{bol}} \sim 10^{35}$ erg s^{-1} in a quiescent,

[152] For tilted disks the sign of a_* is the sign of $\mathbf{J} \cdot \mathbf{L}$, where $\mathbf{J}$ is black hole spin angular momentum and $\mathbf{L}$ is accretion flow orbital angular momentum.

nonflaring state, so that

$$\frac{L_{\rm bol}}{L_{\rm Edd}} = 1.9 \times 10^{-10}\left(\frac{L_{\rm bol}}{10^{35}\ {\rm erg\ s^{-1}}}\right), \quad (8)$$

an extremely small Eddington ratio.

2.2. One-zone Model

Here we motivate the more complicated models that follow using a simple one-zone model, following M87* Paper V and one-zone models developed in the literature over many decades (e.g., Falcke 1996).

Consider a uniform sphere of plasma with radius $r = 5GM/c^2$, comparable to the observed size of Sgr A* at 230 GHz (Paper III; Paper IV), with uniform magnetic field oriented at $\pi/3$ to the line of sight. In turbulent astrophysical plasmas, it is common for the gas pressure to be comparable to the magnetic pressure, so we set $n_i k_{\rm B} T_i + n_e k_{\rm B} T_e = B^2/(8\pi)$, where $T_i \equiv$ ion temperature, $T_e \equiv$ electron temperature, $k_{\rm B} \equiv$ Boltzmann constant, and $B \equiv$ magnetic field strength. The plasma is collisionless (as shown below), and it is plausible that the ions are preferentially heated, so we assume $T_i = 3T_e$. If the ions are subvirial by a factor of 3, as commonly seen in relativistic MHD simulations, i.e., $(3/2)kT_i \sim (1/3)(1/2)(GMm_p/r)$, then the ions are nonrelativistic and the electrons are relativistic, with $\Theta_e \equiv k_{\rm B}T_e/(m_e c^2) \sim 10$.

Assuming a thermal electron distribution function (eDF) and therefore a thermal synchrotron emissivity j_ν (e.g., Leung et al. 2011) and assuming optically thin emission, the flux density from a uniform sphere $F_\nu = (4/3)\pi r^3 j_\nu D^{-2} 10^{23}$ Jy. Requiring $F_\nu = 2.4$ Jy, the average measured by ALMA during the 2017 campaign (Wielgus et al. 2022) yields a nonlinear equation for the electron density n_e with solution

$$n_e \simeq 1.0 \times 10^6\ {\rm cm^{-3}}, \quad (9)$$

$$B \simeq 29\ {\rm G}. \quad (10)$$

This is consistent with a similar one-zone model fit to archival Sgr A* millimeter spectra (Bower et al. 2019). The synchrotron optical depth $\tau_{\rm sync} = r j_\nu / B_\nu \simeq 0.4$, where B_ν is the Planck function, so the optically thin approximation is marginal.

The one-zone model has electron-scattering optical depth $\tau_e = \sigma_{\rm T} n_e r \simeq 2 \times 10^{-6}$, and thus the Compton parameter $y = 16\Theta_e^2 \max(\tau_e, \tau_e^2) \simeq 0.003$ is small. Synchrotron cooling therefore dominates Compton cooling.

The synchrotron cooling timescale for electrons $t_{\rm cool} \equiv u_e/\Lambda$, where $u_e = 3n_e kT_e$ is the electron internal energy and $\Lambda \simeq 5.4 B^2 e^4 n_e \Theta_e^2/(c^3 m_e^2)$ is the synchrotron cooling rate for a thermal population of electrons with $\Theta_e \gtrsim 1$ (see Appendix A in Mościbrodzka et al. 2011; finite optical depth reduces Λ). Therefore, $t_{\rm cool} = 2.3 \times 10^4\ {\rm s} \simeq 1.1 \times 10^3 GM/c^3$, which is longer than the inflow timescale $r/v^r \sim r^{3/2}$. This suggests that radiative cooling can be neglected in the plasma models.[153] More detailed calculations confirm this estimate (Chael et al. 2018; Yoon et al. 2020).[154]

The one-zone model solution implies that the mean free path to Coulomb scattering is large compared to GM/c^2, i.e., the source plasma is collisionless. At $\Theta_e \sim 1$, for example, the electron−electron Coulomb scattering cross section is comparable to the Thomson cross section, and the mean free path is therefore $\sim \tau_e^{-1} GM/c^2$. The electron−ion Coulomb scattering mean free path is even longer, and the electrons and ions are therefore poorly coupled. This is consistent with our assumption that the ions and electrons can have different temperatures (Shapiro et al. 1976; Ichimaru 1977; Rees et al. 1982) and motivates consideration of nonthermal (unrelaxed) eDFs (see Özel et al. 2000; Chan et al. 2009; Mościbrodzka et al. 2014; Davelaar et al. 2018; Chatterjee et al. 2021; Cruz-Osorio et al. 2021; El Mellah et al. 2021; Scepi et al. 2022; Fromm et al. 2022).

2.3. Numerical Models

The one-zone model is too simple for comparison with EHT data. In particular, it does not predict EHT image morphology, and it fails to model emission that arises outside the near-horizon region, including 86 GHz emission and X-ray emission. Steady spherical accretion models (e.g., Falcke et al. 2000) go one step beyond the one-zone model, incorporating relativistic gravity and a radially extended flow. Steady, disk-like (RIAF) accretion models in the Kerr metric go still further and include rotation and departures from spherical symmetry (e.g., Broderick et al. 2009; Huang et al. 2009; Pu & Broderick 2018). Steady phenomenological models do not, however, self-consistently capture fluctuations in the flow. That requires either a statistical model (Lee & Gammie 2021) or a time-dependent numerical simulation. Here we use numerical simulations, adopt an ideal GRMHD model for the flow, employ simple parameterized models to assign an eDF, and solve the radiative transfer equation along geodesics to produce simulated images.

2.3.1. Plasma Flow Model

We model the plasma flow around Sgr A* using ideal, nonradiative GRMHD in the Kerr metric, with a_* a free parameter (see, e.g., Koide et al. 1999; Komissarov 2001; Gammie et al. 2003; De Villiers & Hawley 2003; Anninos et al. 2005; Del Zanna et al. 2007).

We integrate the GRMHD equations in three spatial dimensions using multiple algorithms: KHARMA (Prather et al. 2021), `BHAC` (Porth et al. 2017; Olivares et al. 2019), `H-AMR` (Liska et al. 2019), `koral` (Sądowski et al. 2013), and `Athena++` (White et al. 2016); see Porth et al. (2019) and H. Olivares et al. (2022 in preparation) for comparisons of GRMHD codes. All simulations assume constant adiabatic index $\Gamma_{\rm ad}$.

Unless stated otherwise, the initial conditions for the GRMHD simulations are constant angular momentum hydrodynamic equilibrium tori (Fishbone & Moncrief 1976), with orbital angular momentum that is parallel or antiparallel to the black hole spin. The tori are seeded with a weak, poloidal magnetic field. The simulations use varying torus pressure maximum radius (from $\sim 15\, GM/c^2$ to $40\, GM/c^2$), peak temperature, adiabatic index, and initial field configurations. These variations permit us to test the robustness of our results (see Appendix A).

The torus initial conditions are motivated by the notion that the near-horizon flow, where most of the emission is generated (M87* Paper V), relaxes to a statistically steady state that is

[153] The cooling time for 2.2 μm emitting electrons is $\sim 60(B/(30{\rm G}))^{-3/2}\, GM/c^3$, so cooling is a more significant source of uncertainty for 2.2 μm emission.

[154] If Sgr A* is fed by stellar winds, then the inflowing plasma may be mainly helium (Ressler et al. 2019); this changes the one-zone model slightly. Helium accretion is discussed in G. N. Wong & C. F. Gammie (2022, in preparation).

nearly independent of the flow at larger radius. This notion is challenged in the stellar-wind-fed models of Ressler et al. (2020b), which are included in our study.

All simulations are run in Kerr−Schild-like coordinates, which are regular on the horizon. Unless stated otherwise, boundary conditions are outflow at the inner boundary, which is located inside the event horizon, and outflow at the outer boundary, which is located at $r_{\rm max} \geqslant 1000\,GM/c^2$. Most simulations are evolved to $t_{\rm final} \geqslant 30{,}000\,GM/c^3$.

Once the evolution has started, the magnetorotational instability (MRI; Balbus & Hawley 1992) and possibly other instabilities, such as, for MAD models, magnetic Rayleigh–Taylor instabilities (Marshall et al. 2018), drive the torus to a turbulent, fluctuating state. Defining $P_{\rm gas} \equiv$ gas pressure and $P_{\rm mag} \equiv B^2/(8\pi) \equiv$ magnetic pressure, the standard accretion flow models can be divided by latitude into three zones: (i) an equatorial inflow, (ii) a midlatitude disk wind/corona with $\beta \equiv P_{\rm gas}/P_{\rm mag} \sim 1$, and (iii) a polar "funnel" with $\sigma \equiv B^2/(4\pi\rho c^2) \gg 1$.

The magnetic flux through the horizon, characterized by $\phi \equiv \Phi_{\rm BH}(\dot{M} r_{\rm g}^2 c)^{-1/2}$ ($\Phi_{\rm BH} \equiv$ magnetic flux interior to the black hole equator, $\dot{M} \equiv$ mass accretion rate), divides the outcome into two states: the MAD state (e.g., Bisnovatyi-Kogan & Ruzmaikin 1974; Igumenshchev et al. 2003; Narayan et al. 2003; Tchekhovskoy et al. 2011) and the SANE state (e.g., De Villiers et al. 2003; Gammie et al. 2003; Narayan et al. 2012). MAD models have $\phi \sim \phi_{\rm crit} \sim 60$.[155] In MAD models, magnetic flux accretes onto the hole until $\phi \gtrsim \phi_{\rm crit}$. Accretion of additional flux leads to flux expulsion events so that the flow maintains $\phi \sim \phi_{\rm crit}$. Our SANE models, in contrast, typically have $\phi \sim 1$.

We consider two GRMHD simulations with initial conditions that differ from the fiducial aligned torus: strongly magnetized non-MAD tilted torus simulations (Liska et al. 2018; Chatterjee et al. 2020), and a simulation in which Sgr A* is fed directly by winds from stars in its vicinity (Ressler et al. 2020b). The wind-fed simulations result in a mode of accretion that is similar to MAD but typically has lower mean angular momentum and is less well organized. The wind-fed models have $a_* = 0$.

The GRMHD simulation library is summarized in Table 1. Figure 1 shows a few examples of GRMHD simulations for an aligned SANE, an aligned MAD, a tilted torus, and a wind-fed simulation. These simulations vary in numerical method and in numerical resolution. We present more information on numerical methods in Appendices A and B.

The gas temperature profile is a critical feature of the GRMHD simulations. Figure 2 shows the time- and azimuth-averaged profiles of the midplane dimensionless gas temperature $P/(\rho c^2)$ in a set of aligned GRMHD simulations. The temperature profiles exhibit trends with spin and magnetic state (MAD or SANE) that drive many of the trends seen in the radiative models: MAD models are a factor of several hotter than SANE models, and both MAD and SANE become hotter as a_* increases.

2.3.2. Radiative Transfer Model

Synthetic images are generated from the GRMHD simulations in a radiative transfer step. The transfer step requires (i) a model for the eDF, (ii) assignment of a density scale to the GRMHD simulation, (iii) the inclination i (angle between the torus angular momentum and the line of sight), and (iv) a numerical integration performed as a post-processing step that assumes that the plasma evolution is unaffected by radiation.

Electron Distribution Function.—Thermal models have electron energies distributed according to the Maxwell–Jüttner distribution function:

$$\frac{1}{n_e}\frac{dn_e}{d\gamma} = \frac{\gamma^2\sqrt{1-1/\gamma^2}}{\Theta_e K_2(1/\Theta_e)} \exp\left(-\frac{\gamma}{\Theta_e}\right), \tag{11}$$

where K_2 is a modified Bessel function of the second kind and γ is the electron Lorentz factor. Recall $\Theta_e = k_{\rm B}T_e/(m_e c^2)$, which is determined by the ion−electron temperature ratio $R \equiv T_i/T_e$:

$$T_e = \frac{2m_p u}{3k_{\rm B}\rho(2+R)}. \tag{12}$$

Here u and ρ are the internal energy density and rest-mass density from the GRMHD simulation, respectively, and we have assumed that the ions are nonrelativistic with adiabatic index $5/3$ and the electrons are relativistic with adiabatic index $4/3$. Thermal models are motivated by the idea that wave-particle scattering drives partial relaxation of the eDF, even though Coulomb scattering is ineffective.

The temperature ratio depends on a balance between microphysical dissipation, radiative cooling, and fluid transport. Models for collisionless dissipation vary widely in their predictions for the ratio of heat deposited in ions and electrons but depend most strongly on the local magnetic field strength. This motivates a prescription in which the temperature ratio depends solely on the plasma $\beta \equiv P_{\rm gas}/P_{\rm mag}$ (Chan et al. 2015b). We adopt the same model as M87* Paper V and Event Horizon Telescope Collaboration et al. 2021a, hereafter M87* Paper VIII, where

$$R = \frac{T_i}{T_e} = R_{\rm high}\frac{b^2}{b^2+1} + R_{\rm low}\frac{1}{b^2+1} \tag{13}$$

(Mościbrodzka et al. 2016) and $b \equiv \beta/\beta_{\rm crit}$. This model has three free parameters: $\beta_{\rm crit}$, $R_{\rm low}$, and $R_{\rm high}$. We fix $R_{\rm low} = 1$ (consistent with the long cooling time in Sgr A*; see discussion in Event Horizon Telescope Collaboration et al. 2021a) and $\beta_{\rm crit} = 1$, but we allow $R_{\rm high}$ to vary from 1 to 160. When $R_{\rm high} \gg 1$, emission is shifted away from the midplane and toward the poles.

In *nonthermal* models, the eDF has a power-law tail extending to high energy. We explore two implementations: (i) a power-law distribution function

$$\frac{1}{n_e}\frac{dn_e}{d\gamma} = \frac{p-1}{\gamma_{\rm min}^{1-p} - \gamma_{\rm max}^{1-p}}\gamma^{-p}, \tag{14}$$

with power-law index p and upper and lower limits $\gamma_{\rm min}$ and $\gamma_{\rm max}$; and (ii) a so-called κ distribution function, inspired by observations of the solar wind and by results of collisionless plasma simulations (e.g., Kunz et al. 2015, and references therein)

$$\frac{1}{n_e}\frac{dn_e}{d\gamma} = \gamma\sqrt{\gamma^2-1}\left(1+\frac{\gamma-1}{\kappa w}\right)^{-(\kappa+1)}, \tag{15}$$

[155] In the Lorentz−Heaviside units commonly used in GRMHD simulations $\phi_{\rm crit}$ is smaller by a factor of $(4\pi)^{1/2} \simeq 3.545$.

which has width parameter w and power-law index parameter κ.

Evidently, any eDF assignment scheme is an approximation since the eDF depends in general on both local conditions and particle histories. Notice that we also assume that the eDF is isotropic and neglect electron−positron pairs.

Once the eDF is specified, the radiative transfer coefficients (emissivities, absorptivities, and rotativities) can be readily calculated; see Marszewski et al. (2021) for a recent summary.

Model Scaling.—With the exception of the stellar-wind-fed simulations, the GRMHD simulations considered in this work contain a characteristic speed, c, but are otherwise scale-free; they set $GM = c = 1$. Physical scales are assigned during the radiative transfer step. The black hole mass M fixes the length unit GM/c^2 and time unit GM/c^3. Because the GRMHD simulations are non-self-gravitating, one is free to set a density scale, or equivalently the accretion rate $\dot{M}$ or plasma mass scale $\mathcal{M}$.

The plasma mass scale parameter $\mathcal{M}$ controls the plasma emissivity and the plasma optical depth and thus the source brightness. We adjust $\mathcal{M}$ iteratively until the time-averaged 230 GHz flux densities of the models are within a few percent of the 2.4 Jy mean observed during the 2017 campaign. Notice that, in this work, model parameters are always varied at constant time-averaged millimeter flux density.

Radiative Transfer Calculation.—Given an eDF, density scale $\mathcal{M}$, inclination i, and radiative transfer coefficients based on local properties of the plasma, the emergent radiation is obtained by integrating the radiative transfer equation. We use two classes of numerical methods: observer-to-emitter ray-tracing to generate synthetic images (`ipole`, Mościbrodzka & Gammie 2018; `BHOSS`, Younsi et al. 2012), and emitter-to-observer Monte Carlo to generate spectral energy distributions (SEDs, using `grmonty`; Dolence et al. 2009).

All radiative transport calculations are carried out using the fast-light approximation, in which plasma variables are read from a GRMHD output file at constant Kerr−Schild time and are assumed not to change during ray-tracing. Including light-travel time effects in the model introduces minor changes to light curves and images (Dexter et al. 2010; Mościbrodzka et al. 2021). Further detail on numerical methods is given in Appendix A.1. Comparisons of radiative transfer codes (Gold et al. 2020; B. Prather et al. 2022, in preparation) show that differences between codes do not contribute substantially to the error budget.

Images are produced at 86 GHz, 230 GHz, and 2.2 μm. Direct imaging includes synchrotron and bremsstrahlung (both ion−electron and electron−electron; see Yarza et al. 2020, for a recent review). Unless stated otherwise, the image library has a field of view (full width), resolution (pixel count), and half-width angular size of 800 μas, 200×200, $80\vartheta_g$ at 86 GHz; 200 μas, 400×400, $20\vartheta_g$ at 230 GHz; and 100 μas, 200×200, $10\vartheta_g$ at 2.2 μm.

SEDs are produced for narrow bins in inclination angle. At each inclination, the SED is averaged over azimuth. The SED includes synchrotron, bremsstrahlung, and Compton scattering.

We find that 2.2 μm emission is usually dominated by synchrotron, but occasionally 2.2 μm synchrotron is so weak that Compton scattering dominates. We also find that the X-ray can be dominated by either Compton scattering or bremsstrahlung, with the latter dominating in models with a large population of cold electrons at large radius. Figures 3 and 4 show examples of model images and multiwavelength SEDs from our library.

The GRMHD simulation-derived temperatures are unreliable in regions where $\sigma \equiv B^2/(8\pi\rho c^2)$ is large because truncation error in integration of the total energy equation produces large fractional errors in temperature. All radiative transfer models therefore set the emissivity, absorptivity, and inverse Compton scattering cross sections to 0 for the regions with $\sigma > \sigma_{\rm cut} = 1$.

2.4. Summary of Sgr A* Model Library

A summary of radiative transfer calculations is given in Table 2. The entire image library contains six simulation sets, ∼1.8 million images at each of 86 GHz, 230 GHz, and 2.2 μm, and ∼1.3 million SEDs. The images and SEDs together occupy about 50 TB.

We refer to the thermal, $R_{\rm high}$ models as "fiducial" models and the remainder as "exploratory" models that test the effect of incorporating changes in the eDF or initial conditions. Nearly all exploratory models (exceptions are described in the discussion) are imaged over $5 \times 10^3 GM/c^3$, in comparison to $\geqslant 10^4 GM/c^3$ for the fiducial models. The sampling noise in the exploratory models is therefore larger than in the fiducial models, and thus they cannot be tested as rigorously.

The library contains multiple, redundant models for the fiducial models and variable-κ models. This provides some control over the systematic uncertainties associated with variations in GRMHD simulation setup and algorithms.

3. Observational Constraints

Sgr A* is one of the most frequently observed objects on the sky: it has been observed with a slew of telescopes over five decades in time and 17 decades in electromagnetic frequency. We must select a manageable subset of these data to constrain our models. In doing so, we have attempted to identify constraints (i) that are believed to be uncorrelated, so that each tests a distinct aspect of the model; (ii) that use data that can be simulated with the models; (iii) that are based on EHT 2017 230 GHz VLBI data or that are based on emission produced within or close to the 230 GHz emitting region; and (iv) that are observed contemporaneously or near contemporaneously with the EHT 2017 campaign.

The selected constraints are described in detail below. In brief, the 11 constraints can be divided into three classes. The first class uses EHT data and compares estimates of source size, morphology of the visibility amplitude (VA) distribution, and three parameters of the best-fit m-ring image model (five constraints). The second class uses non-EHT data, including 86 GHz flux density and source size, the median 2.2 μm flux density, and the X-ray luminosity (four constraints). The third class considers variability, including the 230 GHz source-integrated variability and the VA variability based on EHT data (two constraints).

The selected constraints are heterogeneous, and it is not yet possible to combine them in a consistent, fully satisfactory way. Indeed, uncertainties in the data and the models are not well enough understood to make that possible. In this first analysis we set a pass/fail criterion for each constraint and consider the implications of various combinations of constraints.

As the number of constraints increases, so does the probability of wrongly rejecting a model. Consider a set of N

constraints, and for each assign a probability p that the model is consistent with the data. The model is rejected if $p < p_c$. Then, the probability that the model is wrongly rejected by a single constraint is p_c. Applying all N constraints, the probability that the model is wrongly rejected is $1 - (1 - p_c)^N$; for $N = 11$ and $p_c = 0.01$, this is $\approx Np_c \simeq 10\%$. Each of N constraints must therefore be able to reject a model with probability $\ll 1/N$, or the model scoring is meaningless.

The confidence with which a model can be evaluated is limited by sampling noise. Many constraints (e.g., 86 GHz flux density) compare an observation to a distribution of synthetic observations from a model. Time series of synthetic observations are not yet well characterized, but most have a correlation time $\tau \sim$ few $\times\, 100\, GM/c^3$. If the model decorrelates on timescales longer than τ, then a model of duration T yields $\sim T/\tau$ independent samples,[156] and thus a fractional error in moments of the distribution $\sim (T/\tau)^{-1/2} = 0.18(T/15{,}000)^{-1/2}(\tau/500)^{1/2}$. Increasing the number of constraints, then, requires increasing the duration of the GRMHD simulations.

Evidently the models have significant sampling noise, which we control for in part by using three redundant fiducial models. Nevertheless, one should not attach too much significance to the success or failure of individual models.

3.1. EHT Observational Constraints

We test the models against EHT interferometric data in three ways. First, we compare an estimate of the source size ("second moment") against an estimate based on short-baseline VAs. Second, we check the location of the first minimum and the long-baseline values of the VAs ("VA morphology"). Finally, using a variant of a procedure from Paper IV, we compare fits for the diameter, width, and asymmetry of an m-ring (a parameterized image-plane model, "m-ring constraints") to distributions based on synthetic data generated from the model library.

3.1.1. 230 GHz VLBI Pre-image Size

The source size can be characterized using the second moments of the source image on the sky. The second moments in the image domain map to second derivatives of the visibilities near zero baseline in the (u, v) domain, so short-baseline VAs can be used to directly estimate the source size.

This procedure is used in Paper II to set an upper limit of 95 μas FWHM and lower limit of 38 μas FWHM for the second moment along a direction through the source corresponding to the orientation of the short baselines (SMT-LMT and ALMA-LMT). This is done without any assumption about the structure of the source and is therefore quite permissive.

These limits do not include scattering. The scattering kernel is estimated to have 16.2 μas FWHM along the relevant EHT baselines. To descatter the sky image size, we subtract this value in quadrature, which produces a scattering-corrected 93.6 μas FWHM upper limit and 34.4 μas FWHM lower limit.

To score a model, we evaluate the second-moment tensor for each simulated 230 GHz image and find its eigenvalues $\lambda^2_{\rm maj}/(8\ln 2)$ and $\lambda^2_{\rm min}/(8\ln 2)$, where $\lambda_{\rm maj}$ and $\lambda_{\rm min}$ are the major- and minor-axis FWHM, respectively. The image is deemed compliant if there exists any position angle (PA) for which the second moment would satisfy the size constraints, i.e., it is compliant if for any λ such that $\lambda_{\rm min} \leqslant \lambda \leqslant \lambda_{\rm maj}$, λ lies between the scattering-corrected upper and lower limits. We reject models with compliance fraction <0.01.

3.1.2. 230 GHz VLBI Visibility Amplitude Morphology

The second constraint provides a morphological check on the VAs. We ask two questions of each model snapshot: (i) is the first minimum in the visibilities—"the null"—at about the right place, and (ii) are the long-baseline VAs comparable to the data? The null locations and long-baseline amplitudes are sensitive to the source structure. For example, if the source is a simple, circularly symmetric ring of finite width, then the location of the first minimum depends only on the ring diameter, while the VAs on long baselines depend mainly on ring width. GRMHD models are more complicated, with fluctuations in the null locations and long-baseline amplitudes (e.g., Medeiros et al. 2018; M87* Paper V).

We compare with data from April 7, which have the best (u, v) coverage near the minima in the VAs. The first visibility minimum in both the N–S and E–W directions in the data always occurs between 2.5 and 3.5 Gλ (see Paper II, for details). For the long-baseline interval between 6 and 8 Gλ in the data, the VAs have $\lesssim$4% of the zero-baseline flux. One complication when comparing models to data on long baselines is the effect of interstellar scattering. Diffractive scattering effectively convolves the image with a smooth kernel and can reduce the amplitudes to $\sim$50% of their descattered values in the 6–8 Gλ range; refractive scattering, on the other hand, introduces noise at all baselines of order 0.5%–3%, depending on the characteristics of the scattering screen (Johnson et al. 2018; Psaltis et al. 2018).

To apply this constraint, we compute the VA of each model snapshot along PAs of 0°, 45°, 90°, and 135° (because of Hermitian symmetry we need only consider PAs in the 0°–180° range). We find the first minimum numerically and compute the median VAs between 6 and 8 Gλ. We classify a snapshot as compliant if (i) for at least one PA the first minimum falls between 2.5 and 3.5 Gλ and (ii) at no PA do the median VAs exceed $4\%/50\% = 0.08$ of the zero-baseline flux. We reject models with compliance fraction $<1\%$.

3.1.3. 230 GHz M-ring Fitting

Following Paper IV, we fit an m-ring image-plane model to snapshots from EHT data and from simulated data and then compare the distributions of fit parameters.

The m-ring is a δ function in radius with diameter d multiplied by a truncated (up to $m = 3$; notice that Paper IV truncates at $m = 4$) Fourier series, convolved with a Gaussian of width w. The model also contains a centered Gaussian component, with amplitude and width as free parameters, to absorb large-scale emission and emission interior to the ring.[157] The m-ring model has 10 parameters.[158] We use three of the parameters that are well constrained and physically interpretable: the m-ring diameter d, the m-ring width w (FWHM of the convolving Gaussian), and the $m = 1$ relative amplitude β_1 (the

[156] In what follows we must sometimes estimate how many independent samples are available in a time series. Rather than estimating τ model by model, we uniformly assume $\tau \simeq 500GM/c^3$. The analysis is insensitive to this choice.

[157] In Paper IV this is called an mG-ring.

[158] The 10 parameters are ring diameter, ring width, fraction of the flux in the Gaussian component, width of the Gaussian, and six parameters describing the amplitude and phase of the three Fourier components.

Table 1
EHT GRMHD Simulation Library

Setup	Code	a_*	Mode	$\Gamma_{\rm ad}$	$t_{\rm final}$	$r_{\rm out}$	Resolution
Torus	KHARMA[a]	0, ± 0.5, ± 0.94	MAD/SANE	$\frac{4}{3}/\frac{4}{3}$	30,000	1000	288 × 128 × 128
Torus	BHAC[b]	0, ± 0.5, ± 0.94	MAD/SANE	$\frac{4}{3}/\frac{4}{3}$	30,000	3,333	512 × 192 × 192
Torus	H-AMR[c]	0, ± 0.5, ± 0.94	MAD/SANE	$\frac{13}{9}/\frac{5}{3}$	35,000	1000/200	348/240 × 192 × 192
Torus	koral[d]	0, ± 0.3, ± 0.5, ± 0.7, ± 0.9	MAD	$\frac{13}{9}$	101,000	100,000	288 × 192 × 144
Tilted	H-AMR[e]	0.94	SANE[f]	$\frac{5}{3}$	105,000	100,000	448 × 144 × 240
Wind-fed	Athena++[g]	0	MAD	$\frac{5}{3}$	20,000	2,400	356 × 128 × 128

Notes. Summary of the EHT Sgr A* GRMHD simulation library. The last column is $N_1 \times N_2 \times N_3$, with coordinate x_1 monotonic in radius, x_2 monotonic in colatitude θ, and x_3 proportional to longitude ϕ. The first four entries use aligned torus initial conditions. The last two entries are tilted accretion models and two realizations of the wind-fed accretion model that differ in stellar wind magnetization. Times are given in units of $GM/c^3 = 20.4$ s and radii in units of GM/c^2.
[a] See Prather et al. (2021); KHARMA is a GPU-enabled version of the iharm3d code.
[b] Porth et al. (2017); Olivares et al. (2019); Mizuno et al. (2021); Cruz-Osorio et al. (2021).
[c] Liska et al. (2019)
[d] Narayan et al. (2022).
[e] Chatterjee et al. (2020).
[f] $\phi/\phi_{\rm crit} \simeq 0.8$.
[g] White et al. (2016); Ressler et al. (2020b).

"asymmetry"). For more details about the m-ring model see Section 4.3 of Paper IV.

We fit the m-ring independently to snapshots consisting of 2-minute intervals of EHT data (this averaging interval is consistent with that used in Paper IV). Over these short intervals, we approximate the source as static. Uncertainties in the fitted m-ring parameters are dominated by the limited baseline coverage during these snapshots rather than by calibration uncertainties or thermal noise. Because snapshots that are close in time sample nearly identical baselines, they do not provide additional model constraints.

To compare fitted m-ring parameters from EHT data to those from synthetic data, we select a subset of ten 120 s scans that have detections on more than ten baselines and integration times at all stations > 40 s. The selected scans are as widely separated in time as possible so that they sample distinct baseline coverage, with an average separation of $\simeq 1240\,{\rm s} \simeq 60\,GM/c^3$, which is small compared to the VA correlation time in the models (Georgiev et al. 2022). Note that the selected scans overlap with those found in Farah et al. (2022). Only small changes in model selection were observed if any one scan was removed from the comparison. The data were descattered before fitting, that is, the VAs were divided by the scattering kernel.

The maximum likelihood m-ring parameters for the 10 selected EHT scans are listed in Table 3. Evidently the fit parameters are noisy. The fits for d range from 39 to 84 μas, for w from 9 to 21 μas, and for β_1 from 0.04 to 0.48. The variation in fit parameters could be caused by source variability, thermal noise, or gain variations. In the models the main driver of fit variations is source variability.

For the models, we read in a series of model images, generate synthetic data for each image for each scan at four PAs (0°, 45°, 90°, 135°), and fit m-rings to the synthetic data. This produces a distribution of m-ring parameters for each model.

The synthetic data are generated as follows. A model image $I(x, y)$ is Fourier transformed to complex visibilities $V(u, v)$ with an assumed PA and then sampled on baselines i drawn from the comparison scan, $V_i \equiv V(u_i, v_i)$. Normally distributed thermal noise $\delta V_{{\rm th},i}$ with amplitude based on telescope performance during the scan is added, and multiplicative, normally distributed noise with unit variance N is added to crudely model gain corrections: $\tilde{V}_i = V_i(1 + \epsilon N) + \delta V_{{\rm th},i}$. We set $\epsilon = 0.05$, but no substantial changes in fit parameters were observed for $\epsilon = 0.02$. We then fit to the VAs $|\tilde{V}_i|$ and closure phases.[159]

We sample the model images once per 500 GM/c^3, which is comparable to a correlation time. A model with a 15,000 GM/c^3 imaging window thus produces 30 fits per scan per PA.

In comparing the models to the data, we (i) generate the distribution of fit parameters at each PA; (ii) use a Kolmogorov–Smirnov (K-S) test to compare the distribution of ∼300 synthetic data fits with the distribution of 10 observational fits, and obtain a p-value (what is the probability they are drawn from the same underlying distribution?); (iii) average the p-values over the four sampled PAs (i.e., marginalize over PA; the models do not show a significant PA preference); and (iv) reject the model if $p < 0.01$.

3.2. Non-EHT Constraints

In addition to the EHT data, the SED of Sgr A* is well constrained in Paper II and thus potentially useful for model selection. We limit comparison to three bands: 86 GHz, 2.2 μm, and X-ray.

3.2.1. 86 GHz Flux

The Global Millimeter VLBI Array (GMVA) observed Sgr A* on 2017 April 3, just 3 days ($\simeq 13{,}000\,GM/c^3$) before the EHT campaign. Issaoun et al. (2019) estimate that the compact flux during this observation was $F_{86} = 2.0 \pm 0.2$ Jy (2σ errors).

[159] Maximum likelihood m-ring parameters were found for each scan using the Julia package Comrade.jl (P. Tiede 2022, in preparation) in combination with a differential evolution-based optimizer found in the Julia package Metaheuristics.jl. The set of scripts used for the fits can be found in the GitHub repository https://github.com/ptiede/EHTGRMHDCal.

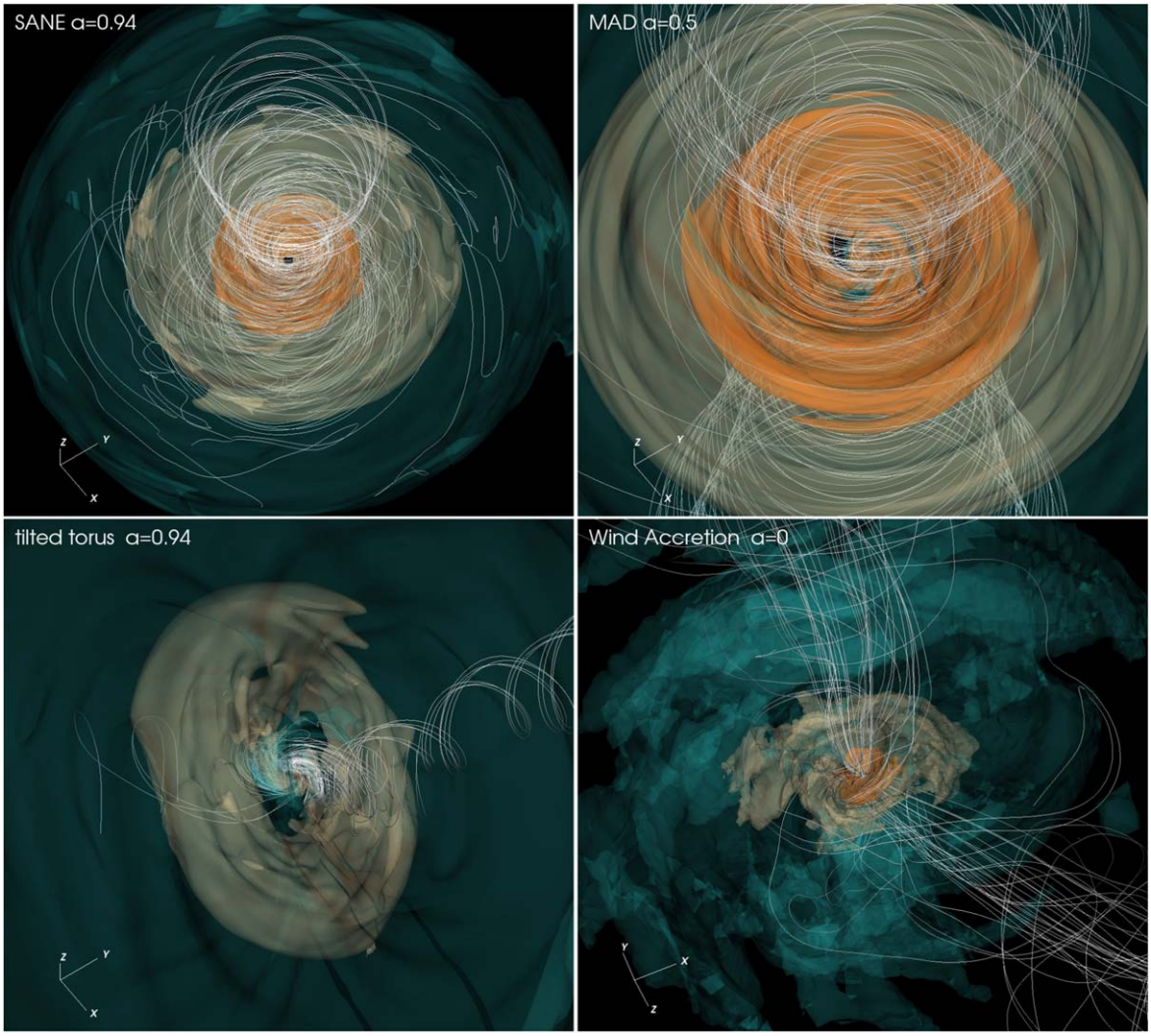

Figure 1. 3D overview of selected GRMHD simulations of Sgr A* in our library. The color marks constant dimensionless density surfaces, and lines follow magnetic field lines. The magnetic field lines shown are only those that are attached to the inner part of the accretion flow, at $r \approx 5\,GM/c^2$. The top panels show accretion simulations with default torus initial condition, and the bottom panels show nonstandard accretion models. The spin is aligned with z-axis.

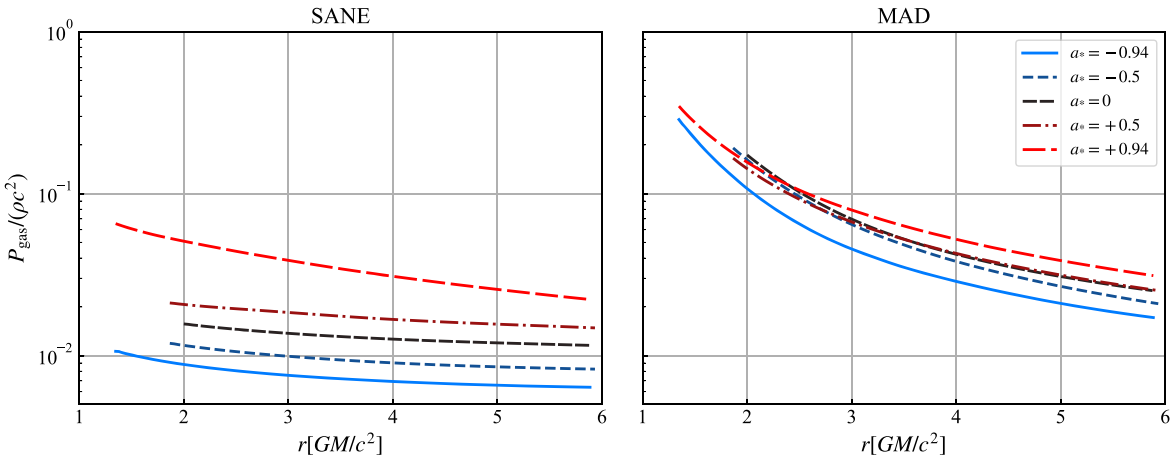

Figure 2. Time- and azimuth-averaged profiles of midplane dimensionless gas temperature $P/(\rho c^2)$ in KHARMA fiducial GRMHD simulations. Evidently MAD models are hotter than SANE, and both MADs and SANEs grow hotter as the black hole spin a_* increases. The hottest models are $a_* = 0.94$ MAD models.

To test the models, we compute a library of 86 GHz images for all GRMHD snapshots for all models and integrate over them to obtain F_{86}. We assume normally distributed measurement errors with $\sigma = 0.1$ Jy and convolve the F_{86} distribution for each model with the resulting Gaussian. We reject models where the value of the error-broadened

cumulative distribution function (CDF) at 2.0 Jy is <1% or >99%.

3.2.2. 86 GHz Image Size

The GMVA observations from 2017 April 3 constrain the FWHM of the source major axis. Notice that two different values for the major-axis FWHM have been published in the literature: $120 \pm 34\ \mu$as (Issaoun et al. 2019) and $\mathrm{FWHM_{maj}} = 146^{+11}_{-12}\ \mu$as (95% confidence; Issaoun et al. 2021). We adopt the latter analysis.

We compute the major-axis FWHM for each image in the 86 GHz image library. We assume normally distributed errors with $\sigma = 6\ \mu$as and convolve the model major-axis distribution with the normal distribution. We reject models with error-broadened CDF < 1% or >99% at 146 μas.

Our synthetic 86 GHz images have a 800 μas field of view. A 200 μas field of view cuts off enough emission that the major axis is biased downward in many models by $\sim 20\%$. Increasing the field of view beyond 800 μas has negligible effect.

3.2.3. 2.2 μm Median Flux Density Constraint

Sgr A* has a quiescent and a flaring component in the near-infrared (NIR), with flares occurring a few times per day (1 day $\simeq 4,200\,GM/c^3$; Witzel et al. 2018). Since there is as yet no generally accepted model for NIR flares, we accept models that do not produce flares (indeed, none of our models reliably produce flares, even those with nonthermal eDFs). Our working hypothesis is that the models can be made to produce flares by introducing a process that accelerates a small fraction of electrons into an NIR-bright tail of the eDF. If the model overproduces quiescent 2.2 μm emission, however, then we reject it.

Sgr A* had a median 2.2 μm flux $= 0.8 \pm 0.3$ mJy in 2017 (Gravity Collaboration et al. 2020a; see Table 1). The median flux density likely overestimates the median quiescent flux density since it includes flares.

We compute the model median 2.2 μm flux density using one of two procedures. If a full SED—which includes Compton scattering—is available, then we use it. The SEDs are generated by the `grmonty` Monte Carlo code (Dolence et al. 2009; Wong et al. 2022). If a full SED is not available (see Table 2), then we compute a 2.2 μm image that includes only synchrotron emission (synchrotron absorption is negligible at 2.2 μm for Sgr A*).

A rigorous model evaluation procedure would correct for the upward bias in median quiescent flux density from flares and allow for errors in the model and observed median flux density, but these refinements are sufficiently uncertain that, instead, we set a conservative threshold of 1.0 mJy and reject the model if its median 2.2 μm flux density exceeds the threshold.

3.2.4. X-Ray Luminosity Constraints

Sgr A* flares in the X-ray less than about once per day (see Yuan et al. 2018, and references therein). Chandra observations during the 2017 campaign suggest a conservative upper limit on the median (quiescent) νL_ν at 6 keV of 10^{33} erg s^{-1} (Paper II).

As for the model 2.2 μm flux density, we estimate $\nu L_\nu|_{h\nu=6\,\mathrm{keV}}$ in two ways. The SED, which incorporates Compton scattering and bremsstrahlung, is used if it is available. If the SED is not available, then we compute an X-ray image that includes only bremsstrahlung (which dominates the X-ray emission in thermal SANE models with $R_{\mathrm{high}} = 40,160$) enabling us to eliminate a few additional models.

We reject the model if its median $\nu L_\nu|_{h\nu=6\,\mathrm{keV}} > 10^{33}$ erg s^{-1}.

3.3. Variability

Sgr A* shows variability on a wide range of timescales. This is expected: fluctuations in stellar wind feeding at the scale of the S stars plausibly introduce long-timescale variations, while turbulence at smaller radii, down to the scale of the event horizon, introduces a spectrum of shorter-timescale variations. Quantitative comparison of observed variability to the models is therefore a potentially powerful tool for model selection.

We consider two variability measures: one characterizes variability in the 230 GHz light curves (Wielgus et al. 2022), and a second characterizes variability of VAs in EHT data (Paper IV; Broderick et al. 2022).

3.3.1. 230 GHz Light Curves

We compare variability in the models to light-curve observations of Sgr A* from 2005 to 2017 using the 3 hr *modulation index* M_3, where $M_{\Delta T} \equiv \sigma_{\Delta T}/\mu_{\Delta T}$, $\sigma_{\Delta T}$ is the standard deviation measured over an interval ΔT (in hours), and $\mu_{\Delta T}$ is the mean measured over the same interval.

Following Chan et al. (2015a), we use $M_{\Delta T}$ because it is easy to describe, easy to compute, commonly used in the literature (in the X-ray astronomy literature it is "rms %"), and closely related to the structure function, since the expectation value for σ_T^2 is given by an integral over the structure function (see D. Lee et al. 2022, in preparation).

We use $\Delta T = 3$ hr ($\sim 530\,GM/c^3$) because it is long enough to be comparable to the characteristic timescale measured in damped random walk fits to the ALMA light curve (see Table 10 of Wielgus et al. 2022) but short enough that the model light curves provide a sample that is large enough to be constraining. In extracting a sample of M_3 from the light curves, we use as many 3 hr segments as possible, equally spaced away from the light-curve endpoints and each other, and calculate M_3 on each segment. We treat consecutive measurements of M_3 as independent, consistent with the minimal correlation expected for a damped random walk (D. Lee et al. 2022, in preparation).

We must select an observed distribution of M_3. The April 7 data alone provide only a weak constraint because there are only three samples. The M_3 measured from EHT 2017 observations on April 5–11 provide seven samples, while the M_3 measured from all available light curves longer than 3 hr, including earlier SMA and CARMA data (the "historical distribution"; see Wielgus et al. 2022), yield 42 samples. The 2017 distribution is consistent with being drawn from the historical distribution, although April 6 has one of the quietest segments on record, and April 11 one of the most variable. We selected the historical distribution and note that the 2017 distribution rejects slightly *more* models but leads to identical conclusions.

For each model we use a two-sample K-S test to estimate the probability p that the model and observed M_3 values are drawn from the same underlying distribution. We reject the model if $p < 0.01$.

Through the K-S test, the strength of the M_3 constraint depends on the number of data and model samples. The fiducial

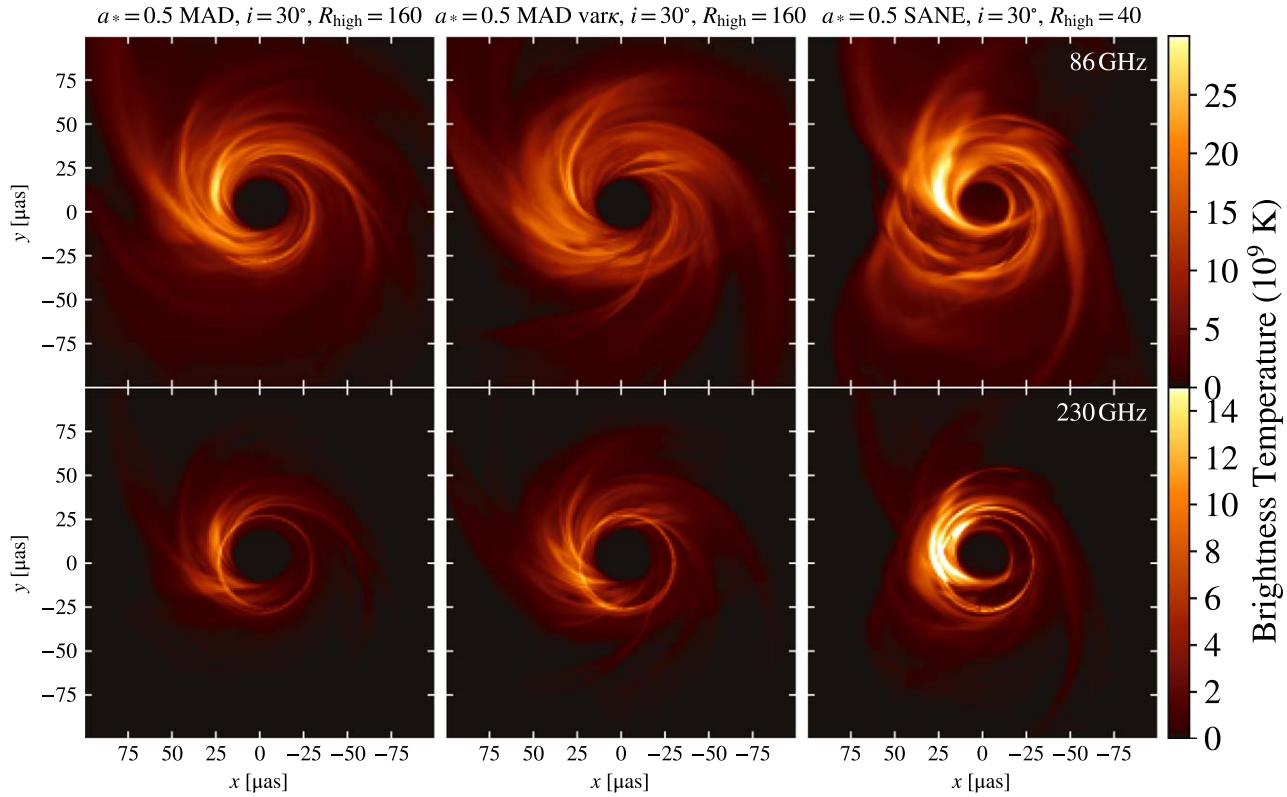

Figure 3. Example images from the model library. Left column: thermal MAD from the best-bet region of parameter space; middle column: nonthermal variable κ MAD; right column: thermal SANE model. Top row: 86 GHz images; bottom row: 230 GHz images. Color represents intensity, or equivalently brightness temperature. Angular momentum of the accretion flow projected onto the image points up. These are relatively successful models satisfying most of the observational constraints.

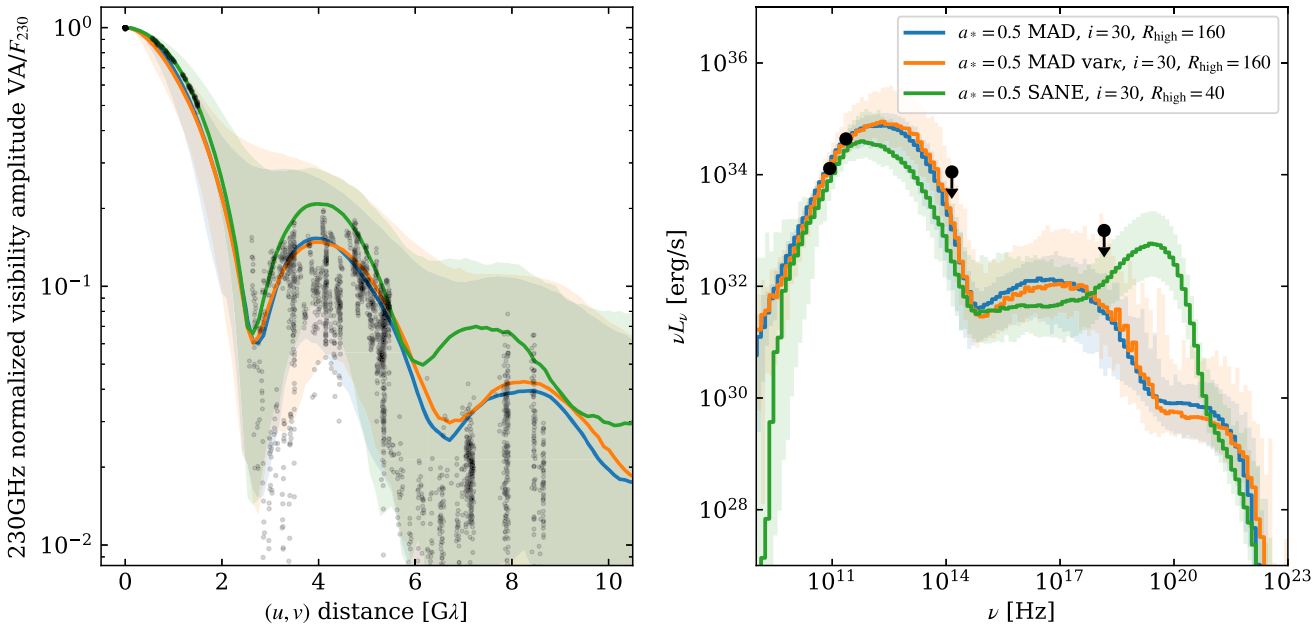

Figure 4. VAs (left) and SEDs (right) of the three example models compared with the calibrated EHT 2017 data. Black symbols show observations. Blue, orange, and green are the models shown in Figure 3. Observed VAs are 1-minute incoherently averaged data from the HOPS pipeline on April 7. Model VAs for a single snapshot are shown as a solid line for a section in the (u, v)-plane at PA = 0°. The band shows the 1st through 99th percentile over all PAs and all times. No noise is included in the model VAs in this figure. Model SEDs (right) show a solid line for the mean SED and a band for the range across snapshots.

Table 2
EHT Model Library

Simulation	Transfer Code	$R_{\rm high}$	Inclination	SED	$\Delta t/(10^3 GM/c^3)$	Notes
			Fiducial Models			
Thermal $R_{\rm high}$ models						
KHARMA	ipole	1, 10, 40, 160	10, 30, ..., 170	Yes	15–30	
BHAC	BHOSS	1, 10, 40, 160	10, 30, ..., 90	Yes	20–30	
H-AMR	BHOSS	1, 40, 160	10, 50, 90	Yes	20–35	
koral	ipole	20	10, 30, ..., 170	No	5–100	
			Exploratory Models			
Thermal $R_{\rm high}$ models						
H-AMR tilted	BHOSS	1, 40, 160	10, 50, 90	Yes	100–103	
Wind accretion	ipole	13, 28	N/A	No	10	
Thermal critical β model						
KHARMA	ipole	N/A	10, 50, 90	No	30–35	
Thermal + power-law models						
H-AMR	BHOSS	1, 40, 160	10, 50, 90	No	30–35	$p = 4$
Thermal + κ models						
BHAC	BHOSS	1, 10, 40, 160	10, 30, ..., 90	No	25–30	$\kappa = 5$
BHAC	BHOSS	1, 10, 40, 160	10, 30, ..., 90	No	25–30	$\kappa = 3.5$ ($\epsilon_0 = 0.05$)
BHAC	BHOSS	1, 10, 40, 160	10, 30, ..., 90	No	25–30	$\kappa = 3.5$ ($\epsilon_0 = 0.10$)
BHAC	BHOSS	1, 10, 40, 160	10, 30, ..., 90	No	25–30	$\kappa = 3.5$ ($\epsilon_0 = 0.20$)
BHAC	BHOSS	1, 10, 40, 80, 160	10, 30, ..., 90	No	25–30	variable $\kappa = \kappa(\beta, \sigma)$
H-AMR	ipole	1, 10, 40, 160	10, 30, ..., 90	Yes	30–35	variable $\kappa = \kappa(\beta, \sigma)$

Note. Summary of the EHT Sgr A* model library. All models are imaged at 86 GHz, 230 GHz, and 2.2 μm, and some (Column (5)) also have SEDs. For the wind-fed accretion model the viewing angle is set by the stellar orbits and $R_{\rm high}$ is set so the model matches the observed 230 GHz flux; $R_{\rm high} = 13$ and 28 for models with weak and strong stellar wind magnetizations, respectively (Ressler et al. 2020b).

models have duration 10^4 or $1.5 \times 10^4\, GM/c^3$ (18 or 28 samples), whereas most exploratory models have duration $5 \times 10^3\, GM/c^3$ (nine samples). The M_3 constraint is therefore weaker for the exploratory models: an exploratory model that passes the constraint may be more variable than a fiducial model that fails.

3.3.2. EHT Structural Variability

Fluctuations in the spatial structure of the source produce fluctuations in the VAs. Here we compare the power spectrum of structural variability from EHT observations with predictions from GRMHD models.

A nonparametric technique to measure the variance of the spatially detrended VAs at a location in the (u, v)-plane is described in Broderick et al. (2022) and briefly summarized here. We use EHT observations of Sgr A* from April 5, 6, 7, and 10 (April 11 was excluded). To remove correlations associated with variations in the total flux, we normalize the VA data with the contemporaneous intrasite light curve (Georgiev et al. 2022). The light-curve-normalized VAs are then linearly detrended, and variances are computed and azimuthally averaged (Broderick et al. 2022). The resulting $\sigma^2_{\rm var}(|u|)$ is a measure of the fractional structural variability as a function of baseline length $|u|$. The $\sigma^2_{\rm var}(|u|)$ is included in an inflated error budget when making images of and fitting models to the 2017 EHT observations of Sgr A* (Paper III).

We measured this quantity from the GRMHD simulations (see Georgiev et al. 2022, for details). For all simulations reported here, $\sigma^2_{\rm var}$ is well approximated by a broken power law with parameters that are nearly universal among simulations. The $\sigma^2_{\rm var}$ is measured over a 4-day period, which is longer than the typical model duration. We therefore expect that model values will be biased downward compared to the data. Furthermore, each GRMHD simulation can only give one draw from a distribution that is broader than if the simulation spanned 4 days. This secondary effect negates the downward bias, which is further unimportant, as we do not exclude models for being not variable enough. To measure the larger broadness of the distribution, we use multiple simulations with the same parameters and subdivide the analysis of long simulations into windows. The uncertainties in the measurement from the GRMHD simulations due to simulation resolution, the fast-light approximation, and code differences are small compared to the uncertainty due to the variability of $\sigma^2_{\rm var}$ due to short simulations (Georgiev et al. 2022).

The measured $\sigma^2_{\rm var}$ is well characterized by a power law for $2\,{\rm G}\lambda < |u| < 6\,{\rm G}\lambda$ (Georgiev et al. 2022). For comparison with the models presented here, we distill the $\sigma^2_{\rm var}$ to two numbers: the amplitude a_4^2 at $4\,{\rm G}\lambda$ and a power-law index b. Because the normalization is done in the center of the fit range, the estimated $\log_{10}(a_4^2) = -3.4 \pm 0.1$ and $b = 2.4 \pm 0.8$ are essentially uncorrelated.

Model predictions for a_4^2 and b are computed using the power spectral densities from Georgiev et al. (2022).[160] The anisotropic diffractive scattering kernel from Johnson et al. (2018) is applied to $\sigma^2_{\rm var}(|u|)$ and averaged over relative orientations of the major axis of the scattering kernel and

[160] Georgiev et al. (2022) gives the power spectral density of the complex visibility, $\langle \hat{P}(|u|) \rangle$, rather than the VA, and thus $\sigma^2_{\rm var} = \langle \hat{P} \rangle / 2$.

the black hole spin. These estimates are then azimuthally averaged, and the parameters a_4^2 and b are determined from a least-squares linear fit to $\sigma^2_{\rm var}(|u|)$ in $2\,{\rm G}\lambda < |u| < 6\,{\rm G}\lambda$.

For each model the fits for a_4^2 and b are done separately on each window of length $5 \times 10^3\, GM/c^3$, giving at most three measurements for most models. This makes a direct comparison with the measured value difficult, as the model distribution is poorly constrained.

Georgiev et al. (2022) estimates that the typical width of a model distribution is $\log_{10}(a_4^2) \pm 0.1$. We can obtain a rough estimate for how the models fare compared to the measurement by taking the mean across windows, assuming that the width of the distribution is $\sigma = 0.1$, and comparing this with the observed distribution under the assumption that both are distributed normally. We reject models with error-broadened CDF <1% or >99% at $\log_{10}(a_4^2) = -3.4$.

4. Model Comparison

4.1. Fiducial Models

We start with the fiducial models. Recall that these have aligned (prograde or retrograde) accretion flows, thermal eDFs, and electron temperature assigned according to the $R_{\rm high}$ model, as in M87* Paper V, and include the KHARMA, BHAC, and H-AMR model sets listed at the top of Table 2.

A set of plots showing how the three, redundant fiducial model sets fare for each constraint is provided in Appendix C. Table 4 summarizes the fraction of fiducial KHARMA, BHAC, and H-AMR models that pass each constraint.

4.1.1. EHT Constraints

Second Moment.—Without assuming a ring, the EHT data allow a wide range of second moments. The second moment constraint passes 98% of all models. Here and in what follows, the quoted passing fraction for the model describes the fraction of points in parameter space for which the existing model sets (KHARMA, BHAC, and when present H-AMR) agree that the model passes the constraint. In short, nearly all fiducial models are about the right size once we use the 230 GHz to fix the mass unit $\mathcal{M}$. The few rejected models are $a_* \leqslant 0$, face-on, SANE models with $R_{\rm high} = 1$. These models have extended emission on scales large compared to the critical impact parameter $b_c = \sqrt{27}\, GM/c^2$. The right panel of Figure 5 shows an example of one of these failed models. The left panel shows an example of a passing model.

Visibility Amplitude Morphology.—The VA morphology constraint tests the null location and long-baseline VAs. Figure 6 shows an example of a passing and failing model. The constraint disfavors edge-on models at positive spin and a few large-$R_{\rm high}$ SANE models. This is mainly because the edge-on models contain bright spots, corresponding to the approaching side of the rotating accretion flow, and faint rings, so the first nulls get washed out by the bright features. The VA morphology constraint passes 79% of all models.

M-ring Fits.—The m-ring asymmetry, diameter, and width are treated as separate constraints. Recall that we compare the distribution from the data to that from the model using a two-sample K-S test.

The asymmetry parameter is typically not well constrained. Many rejected models are at high inclination and have $a_* = 0.94$. These models have asymmetries that are large and detectable because Doppler boosting concentrates emission in an equatorial spot on the approaching side of the disk. The asymmetry parameter constraint passes 91% of all models.

The m-ring diameter, which depends on the diameter of the shadow and the ring width, is better constrained than the asymmetry parameter and varies systematically from model to model. The ring diameter constraint passes 54% of all models.

Most of the models that fail are low-inclination models with ring diameters that are too large. Only two BHAC models fail because the ring diameter is too small. Most of the rejected models are low-inclination models at $a_* < 0$.

The m-ring width w is the most tightly constrained of the three m-ring parameters. Although the closure phases constrain w as well, it is easiest to see how w affects VAs at long baselines. For example, for a circularly symmetric ring the VAs are a Bessel function multiplied by a Gaussian with width $\sim 1/w$. Increasing w therefore decreases the amplitude of the long baselines. Figure 7 shows examples of models that pass and fail the m-ring width constraint.

Figure 8 summarizes the pass/fail status of the fiducial models for the m-ring width. All rejected models have median w that is below the median of the data, $\simeq 17.5\ \mu$as. The rejected models include all MAD models at $a_* \leqslant 0$ and all edge-on ($i = 90°$) models in the KHARMA, BHAC, and H-AMR fiducial models. MAD models exhibit a strong trend toward smaller w as i increases. SANE models exhibit a similar but weaker trend. The SANE model images have higher optical depth, broader rings, and more substructure than the MAD models. Their w distributions are typically broad, with mode well below $17.5\ \mu$as. Only for $a_* = 0.94$, where the optical depth is lower owing to higher temperatures in the emitting region, do most of the models exhibit a sharply peaked w distribution centered at $17.5\ \mu$as.

EHT Constraint Summary.—We can combine all EHT constraint cuts with a logical *and* operation. The results are summarized in Figure 9. Evidently EHT data alone are capable of discriminating between models. The edge-on ($i = 90°$) models all fail, with some failing m-ring width, diameter, asymmetry, and the VA morphology constraint. The cuts clearly favor $a_* > 0$ models, with a few exceptions. There are two clusters of models that do not fail any constraints in any models: positive-spin MAD models at low inclination, and positive-spin SANE models, also at low inclination.

4.1.2. Non-EHT Constraints

86 GHz Flux Density.—In a simplified picture Sgr A*'s millimeter flux is produced in a photosphere that decreases in size as frequency increases. Because optical depth is not large at 230 GHz (~0.4 in the one-zone model) and the source structure is complicated (the optical depth varies across the image), the simplified picture is imprecise. Nevertheless, 86 GHz emission is on average produced at larger radius than 230 GHz emission, and the 86 GHz source size is larger than the 230 GHz source size. The ratio of 86 GHz to 230 GHz flux density is therefore sensitive to the radial structure of the source plasma.

Figure 28 records the results of applying this constraint. Most $R_{\rm high} = 1$ models, both MAD and SANE, fail the 86 GHz flux density test. The 86 GHz flux density is quite sensitive to $R_{\rm high}$. For example, SANE $a_* = 0.5$, $i = 70°$, 90° models are too bright at $R_{\rm high} = 1$ and too dim at $R_{\rm high} = 10$. This suggests

Table 3
M-ring Fits to EHT Observations

Scan #	t (UTC hr)	d (μas)	w (μas)	β_1
111	11.28	83.87	8.87	0.122
121	11.78	57.09	13.98	0.220
125	11.92	55.63	16.46	0.132
130	12.35	40.68	19.08	0.039
134	12.62	57.22	17.22	0.368
142	12.92	58.80	17.55	0.208
149	13.28	52.31	21.16	0.278
155	13.75	38.94	18.17	0.482
163	14.05	56.22	19.86	0.470
171	14.38	39.48	17.71	0.408

Note. The m-ring fits to selected 120 s scans from April 7. Column (2) gives UTC in hours for the observation. Columns (3)–(5) give best-fit parameters for the m-ring diameter, width, and asymmetry parameter, respectively.

that there are passing models in between, and that the parameter space is not sampled densely enough. Finally, the 86 GHz flux constraint strongly favors MAD models over SANE models in all three fiducial model sets.

86 GHz Major Axis.—As for the 86 GHz flux, the 86 GHz size is sensitive to optical depth as a function of radius in the source plasma. Figure 29 in Appendix C shows the full results of applying this constraint.

The 86 GHz size is sensitive to inclination. For example, the SANE, $a_* = 0$, $R_{\rm high} = 40$ models are too small at low inclination and too large when seen edge-on, because the edge-on models have prominent limb-brightened jet walls that are visible to 100 μas. The 86 GHz size constraint passes only 58% of models and is therefore one of the tightest constraints.

The physical picture for 86 GHz source size is complicated, as is the extraction of the constraint itself from observations. Notice that (i) two different values for the 86 GHz intrinsic source size have been reported in the literature (see Section 3.2.2), (ii) scattering is 7 times stronger at 86 GHz than at 230 GHz, (iii) scattering must be subtracted accurately to obtain the intrinsic source size, and (iv) the error bars for the 86 GHz source size are narrow and this plays a key role in determining the strength of the constraint.

2.2 μm Median Flux Density.—2.2 μm photons are produced by the synchrotron process from electrons on the high-energy end of the eDF. For the one-zone model with $B = 30$ G and $\Theta_e = 10$, the mean Lorentz factor is $\gamma = 30$ and the synchrotron critical frequency $\nu_{\rm crit} = \gamma^2 eB/(2\pi m_e c) \simeq 80$ GHz. Emission at 2.2 μm is produced by electrons with Lorentz factor $\gamma \simeq 10^3$, so 2.2 μm flux density is sensitive to Θ_e and B. Both increase toward the horizon, and field strength is nearly independent of latitude, so 2.2 μm photons are produced at small radius in regions where Θ_e is highest.

The sensitivity to Θ_e implies that 2.2 μm flux density will be highest for models with higher temperatures. For SANEs the midplane gas temperature, and therefore electron temperature in the $R_{\rm high}$ prescription, increases with a_*, so the highest 2.2 μm flux density is at positive a_*.

The sensitivity to B implies that 2.2 μm flux density will be highest for parameters with stronger fields. B depends on the GRMHD flow configuration and also on the accretion rate, which is fixed by the observed F_{230}, so when all else is equal the 2.2 μm flux density is highest when the accretion rate is largest. The dependence of accretion rate on model parameters is discussed in Section 5.5. In brief, for SANE models the accretion rate declines as a_* increases and $R_{\rm high}$ decreases. For MAD models the accretion rate dependence on a_* and $R_{\rm high}$ is relatively weak.

Finally, the 2.2 μm flux density is also sensitive to inclination. A combination of Doppler boosting and the rapid falloff in emissivity in the NIR means that at large inclination lower-frequency emission from the approaching side of the accretion flow is boosted into the NIR, and thus 2.2 μm flux is higher at high inclination.

Figure 10 shows sample SEDs from our model library, where the left panel is a model that passes the 2.2 μm flux limit and the right panel is a model that fails. Models that pass the 2.2 μm flux limit are shown in Appendix C in Figure 30. The rejected SANE models (7% rejected by all of KHARMA, BHAC, and H-AMR) tend to be at high inclination: their images are dominated by a bright spot on the approaching side of the disk. The rejected MAD models (53%) include nearly all models at $R_{\rm high} = 1$ and $R_{\rm high} = 10$, where Θ_e tends to be larger, and the majority of high-inclination models, where the effect of Doppler boosting is largest.

We find that some models are Compton dominated at 2.2 μm. For example, $a_* = -0.94$ SANE models become optically thin at relatively low frequency as $R_{\rm high}$ goes to 1, and thus synchrotron emission drops off rapidly as frequency increases. When the synchrotron is weak enough, the underlying bump of Comptonized millimeter photons dominates.

X-ray Luminosity.—X-ray production in fiducial models is typically dominated by Compton upscattering of thermal synchrotron photons. In the first Compton bump νL_ν is thus proportional to the y-parameter $y \sim 16\Theta_e^2\tau_e$, where τ_e is a characteristic electron-scattering optical depth and Θ_e is a typical dimensionless electron temperature. At $R_{\rm high} = 1$ the X-ray band lies in the first Compton bump, while at larger $R_{\rm high}$ the bumps move to lower energy because the bulk of the Thomson depth is in the midplane where $\Theta_e \propto 1/R_{\rm high}$.

We find that in a few large-$R_{\rm high}$ SANE models, however, X-ray emission is dominated by bremsstrahlung (synchrotron never dominates the X-ray in thermal models). Bremsstrahlung emissivity $j_{\nu,b} \propto n^2$, so at fixed temperature bremsstrahlung increases rapidly with density. Notice that $j_{\nu,b} \propto \Theta_e^{1/2}$ for $\Theta_e > 1$ and $\Theta_e^{-1/2}$ for $\Theta_e < 1$, so cool disks enhance bremsstrahlung. Bremsstrahlung therefore dominates Compton in models with high density and low temperature, i.e., some models with large $R_{\rm high}$ (see Section 5).

In models with bremsstrahlung-dominated X-ray emission the median radius of emission is $\simeq 20GM/c^2$. Although the models are equilibrated at this radius, the X-ray luminosity may be partially contaminated by emission from unequilibrated plasma at larger radii. Because the fiducial models start with a torus of finite radial extent, however, they are also missing bremsstrahlung emission from outside the initial torus. A full assessment of the associated uncertainty requires large, long runs. Notice that because bremsstrahlung arises at large radii, it varies more slowly than the synchrotron and Compton-upscattered X-ray emission and is therefore potentially distinguishable (Neilsen et al. 2013).

The left panel of Figure 10 shows a model that passes the X-ray flux limit, while the right panel shows a model that fails. The X-ray cuts are shown in Appendix C, Figure 31. Some large-$R_{\rm high}$ SANE models fail owing to excess bremsstrahlung, although there is notable disagreement between BHAC and

KHARMA for SANE X-ray fluxes. MAD models that fail have small $R_{\rm high}$ and are Compton dominated in the X-ray. Nearly all $R_{\rm high} = 1$ MAD models fail the X-ray constraint, as do many at $R_{\rm high} = 10$. This is because the midplane Θ_e increases as $R_{\rm high}$ goes to 1. Since the midplane contributes most of the electron-scattering optical depth, small-$R_{\rm high}$ models have the largest y-parameter and are at greatest risk of overproducing X-rays.

Summary of Non-EHT Constraints.—Applying only non-EHT constraints leaves 6% of models as shown in Figure 11. The surviving models are the result of applying a heterogeneous and noisy set of constraints using a hard cutoff, which somewhat obscures the underlying physical picture. Nevertheless, the surviving 13 models are all MAD and all have $R_{\rm high} > 10$. All but two have $i < 70°$. This leaves a cluster of surviving MAD models at large $R_{\rm high}$ and low to moderate inclination.

4.1.3. Variability

Variability is central to the interpretation of EHT observations of Sgr A*: an 8 hr observation of Sgr A* lasts $1400\,GM/c^3$, a timescale over which most models vary substantially. In contrast, an 8 hr observation of M87* is $\sim GM/c^3$, and on this timescale M87* hardly varies at all.

Recall that we consider two variability constraints, one on the 230 GHz light curve and the other on 230 GHz VAs. We find that SANE models are less variable than MAD models. Only 3.5% of models, all SANE, pass both variability constraints. A possible interpretation of this result is that the models are missing a physical ingredient that would reduce variability, and this is discussed in Section 5.

Modulation Index.—The distributions of 3 hr modulation index (M_3) across all fiducial SANE models, across all fiducial MAD models, and across the historical data set are shown in Figure 12. The plot also shows distributions for individual models with the lowest and highest median M_3.

The M_3 cuts are summarized in Appendix C, Figure 34. We find that (i) as a group, the fiducial models are more variable than the data; (ii) the MAD models are more variable than SANE models; (iii) 11 individual models pass the constraint for all fiducial model sets, and these are exclusively SANE models; (iv) there are some differences between variability in the fiducial model sets, with H−AMR models notably more variable than KHARMA and BHAC models; and (v) the pass fractions for the fiducial model sets are 20% for KHARMA, 27% for BHAC, and 7% for H−AMR. The modulation index is the tightest single constraint on the models.

4 Gλ Visibility Amplitude Variability.—The power-law index b of the variance $\sigma^2_{\rm var}(|u|)$ at 2–6 Gλ of the models is generally in good agreement with the value measured from the 2017 EHT campaign (excluding April 11). The amplitude a_4^2, however, varies depending on the model.

Figure 13 shows the distribution of (a_4^2, b) from the EHT observation, along with the distributions across all fiducial models. For a single model, the number of measurements of (a_4^2, b) is equal to the number of windows for that model (three in most cases). The koral models appear more variable because they include only $R_{\rm high} = 20$ MAD models at various spins.

The models tend to be more variable than the observations, with face-on models performing better than edge-on models. For SANE models, $R_{\rm high} = 10$ tends to be more variable than others. For MAD models, there is a slight preference for lower $R_{\rm high}$.

Long-duration koral Models.—We have imaged a set of MAD models run with the koral code out to $\sim 100,000\,GM/c^3$. These long-duration models have $R_{\rm high} = 20$, which lies off our fiducial model parameter grid. They enable us to assess the importance of integration time for application of the constraints and provide a more accurate distribution for, e.g., M_3.

The koral models are discussed in Appendix B. In brief, we find no evidence for significantly different variability when comparing the first and second half of the koral runs, consistent with no long-term evolution of the variability. We also find no significant differences when comparing the koral runs to nearby models on the fiducial model parameter grid. Notice that in Figure 13 the koral models are more variable than the other model sets only because the other model sets contain lower-variability SANE models.

4.1.4. Summary of Constraints on Fiducial Models

None of the fiducial models survive the full gauntlet of constraints. The pass fractions for individual constraints for the BHAC, KHARMA, and H−AMR fiducial models are listed in Table 4. M_3 is the most severe constraint, followed by the m-ring width constraint. Together the variability constraints pass only 4% of fiducial models and prefer SANEs, which are less variable than MADs, while the remaining constraints prefer MAD models.

It is likely that the models are physically incomplete. It is also possible, however, that one of the constraints is measured incorrectly, that one of the constraints is applied incorrectly, or that one of the constraints is poorly predicted for numerical reasons. To investigate this, we identify all models that fail only one constraint in Table 5. We find that the critical constraints are 86 GHz size, m-ring diameter, and M_3. Notice that there is overlap between KHARMA and BHAC in MAD models that fail the M_3 constraint. The H−AMR models fare significantly worse than the KHARMA and BHAC models in the M_3 constraint: only 7% of models pass. The remaining models all fail at least one additional constraint, leading to their exclusion from Table 5.

4.2. Exploratory Models

Next, we go beyond the fiducial models and consider the exploratory models, which include aligned models that use an alternative scheme for assigning temperatures to a thermal eDF, aligned models with a power-law component or κ component in the eDF, tilted models, and stellar-wind-fed models. Unless stated otherwise, exploratory models are imaged over only $5 \times 10^3\,GM/c^3$, yielding weaker constraints. In all cases we focus on how the exploratory models differ from the fiducial models.

4.2.1. Critical Beta Model

The $R_{\rm high}$ prescription provides a convenient, one-parameter model for assigning electron temperatures, but here is a vast function space of possible alternative parameterizations. One well-motivated choice is the critical beta model (Anantua et al. 2020b), which sets $T_e = T_e(R)$ and $R = f\exp(-\beta/\beta_c)$ (see Equation (11)). This "critical beta" model has two parameters, f and β_c. We consider a single point in the parameter space:

Table 4
Fiducial Model Pass Fractions

Constraint	KHARMA	BHAC	H-AMR
230 GHz size	0.98	0.98	1.0
VA morphology	0.84	0.83	0.80
M-ring diameter	0.67	0.65	0.58
M-ring width	0.35	0.21	0.29
M-ring asym.	0.94	0.95	1.0
86 GHz flux	0.74	0.68	0.62
86 GHz size	0.65	0.59	0.46
2.2 μm flux	0.59	0.55	0.80
X-ray flux	0.46	0.70	0.61
Light-curve variability	0.20	0.27	0.07
4 Gλ variability	0.60	0.72	0.39
EHT constraints	0.25	0.19	0.22
Non-EHT constraints	0.19	0.19	0.22
Variability constraints	0.16	0.27	0.03

Note. Passing fractions for the fiducial KHARMA, BHAC, and H-AMR $R_{\rm high}$ thermal models, showing the consistency and relative power of the constraints.

$f = 0.5$, $\beta_c = 1$. Compared to the $R_{\rm high}$ temperature prescription, the main new characteristic of the critical beta models is that the electron-to-ion temperature ratio approaches 0 at high β instead of $1/R_{\rm high}$.

We have run all tests except X-ray for the critical beta models. The 2.2 μm flux is calculated by imaging only and therefore does not include Compton scattering.

All critical beta models fail the non-EHT constraints, with the 86 GHz size constraint rejecting most models as too small. The variability constraints pass 23% of the models. No models survive the combined EHT and non-EHT constraints even if variability constraints are excluded. Notice that this does not imply that critical beta models are ruled out, since we have only tested a single point in the f, $\beta_{\rm crit}$ parameter space.

4.2.2. *Thermal Plus* p = 4 *Power-law Models*

So far we have assumed a thermal eDF (Equation (11)). Fully kinetic simulations and resistive MHD predict that reconnection in current sheets within the accretion flow and in the jet sheath leads to the acceleration of particles to higher energies, resulting in the emergence of a power-law tail (e.g., Sironi et al. 2021, and references therein). Such acceleration events are thought to be the origin of NIR and X-ray flares detected in Sgr A*. Here we do not address flare mechanisms but seek to constrain the contribution of nonthermal electrons to the quiescent emission of Sgr A*.

Below we assume different forms of the eDF assuming that a fraction of the electron population is accelerated into a nonthermal tail. There are multiple ways of doing this, but we will continue to assume that the eDF depends instantaneously on local conditions and set the accretion rate so that the 230 GHz time-averaged compact flux is 2.4 Jy.

First, we consider a hybrid thermal/power-law distribution using H-AMR/BHOSS. Since we are modeling quiescent emission, we assume a steep power-law index of $p = 4$ with a constant nonthermal acceleration efficiency $\epsilon = n_{e,\ \text{power-law}}/n_{e,\ \text{thermal}} = 0.1$, typical of PIC simulations (e.g., Sironi et al. 2015; Crumley et al. 2019). Following Chatterjee et al. (2021), the power-law tail is stitched to the thermal core by choosing the minimum Lorentz factor limit of the power law, $\gamma_{\rm min}$, to be at the peak of the Maxwellian component. The upper end of the power law is set to $10^5\gamma_{\rm min}$ (see Equation (14)). The temperature of the thermal component is set by the $R_{\rm high}$ prescription (Equation (13)). We find that the accretion rate is slightly smaller than for corresponding thermal models, consistent with a small contribution from the power-law component to the 230 GHz total intensity.

230 GHz VLBI Pre-image Size.—Hybrid thermal/power-law models have larger 230 GHz VLBI pre-image sizes compared to their purely thermal counterparts. This is because the power-law component of the eDF allows high-energy electrons in weak magnetic fields at distances more than a few gravitational radii (i.e., larger than the typical emission radius of the 230 GHz images) to contribute to the total image. However, the extension in the images is much smaller for MAD models, with most MAD images displaying an increase in size of $<10\%$.

86 GHz Flux and Image Size.—In general, the $R_{\rm high} = 1$ models produce too much 86 GHz flux. Since the lower limit of the power law $\gamma_{\rm min}$ is directly affected by the local electron temperature, the highest-energy electrons are located in the jet sheath where $T_i \approx T_e$. Indeed, this is why SANE models produce more 86 GHz flux when nonthermal electrons are introduced, especially at larger $R_{\rm high}$ values. On the other hand, MAD thermal and mixed thermal/nonthermal models behave similarly, as the bulk of the emission is produced in the inner disk.

The 86 GHz image sizes for the hybrid H-AMR models are, on average, larger than their thermal-only counterparts, similar to the 230 GHz image sizes. The higher-energy electrons of a hybrid thermal/power-law population emit at higher frequencies than their thermal core, thereby extending the image size. This effect increases the image size of MAD models by only a few percent.

2.2 μm Constraint.—The addition of the power-law tail increases the flux at 2.2 μm, and thus the GRAVITY-based 2.2 μm median flux density threshold of 1.0 mJy provides a strong constraint on the power-law index and the acceleration efficiency. In brief, 59% of the power-law models, especially $R_{\rm high} = 1$ and 40 MAD models, are ruled out by the 2.2 μm constraint.

Summary.—Overall, H-AMR hybrid thermal/power-law models behave quite differently from their thermal counterparts. For the thermal models, both EHT and non-EHT constraints are equally successful in ruling out models, with 22% passing for each constraint set. For the power-law model set non-EHT constraints pass 39% of models while EHT constraints pass 10% of models. This disparity occurs for two reasons: (i) introducing nonthermal electrons pushes the 86 GHz image size to the acceptable range, as thermal models typically exhibit small image sizes; and (ii) the m-ring width is found to be smaller for the hybrid models. This could be due to a change in the gas density scaling that is required to match the 230 GHz flux. Nonthermal models require a smaller normalization value, meaning a smaller electron number density as compared to the corresponding thermal models. A decrease in the number density lowers the optical depth, leading to a thinner photon ring. For the initial $5000\,GM/c^3$ survey, two mid-inclination power-law models survive: a SANE $a_* = 0.94$ model and a MAD $a_* = 0.5$, $R_{\rm high} = 1$ model (see Table 6),

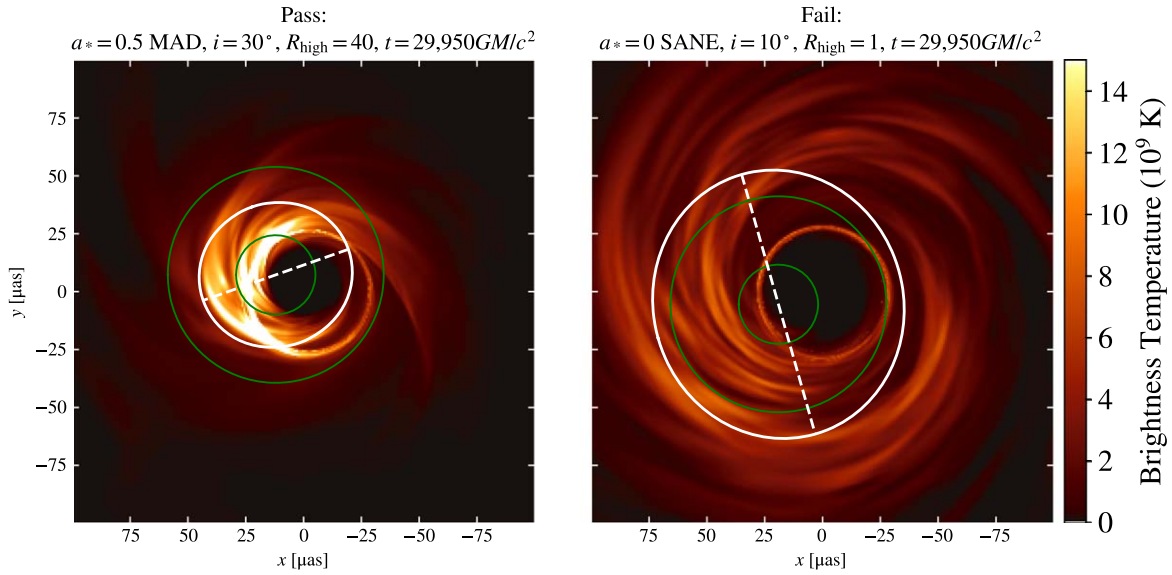

Figure 5. Pre-image size constraint example. Left: passing snapshot; right: failing snapshot. The model is rejected if <1% of model snapshots pass. The solid white ellipse represents the second moments of the image, and the dashed line shows the major axis. The two green circles show the observed lower and upper limits from Paper III. The snapshot is rejected if the minor axis is larger than the upper limit, or if the major axis is smaller than the lower limit.

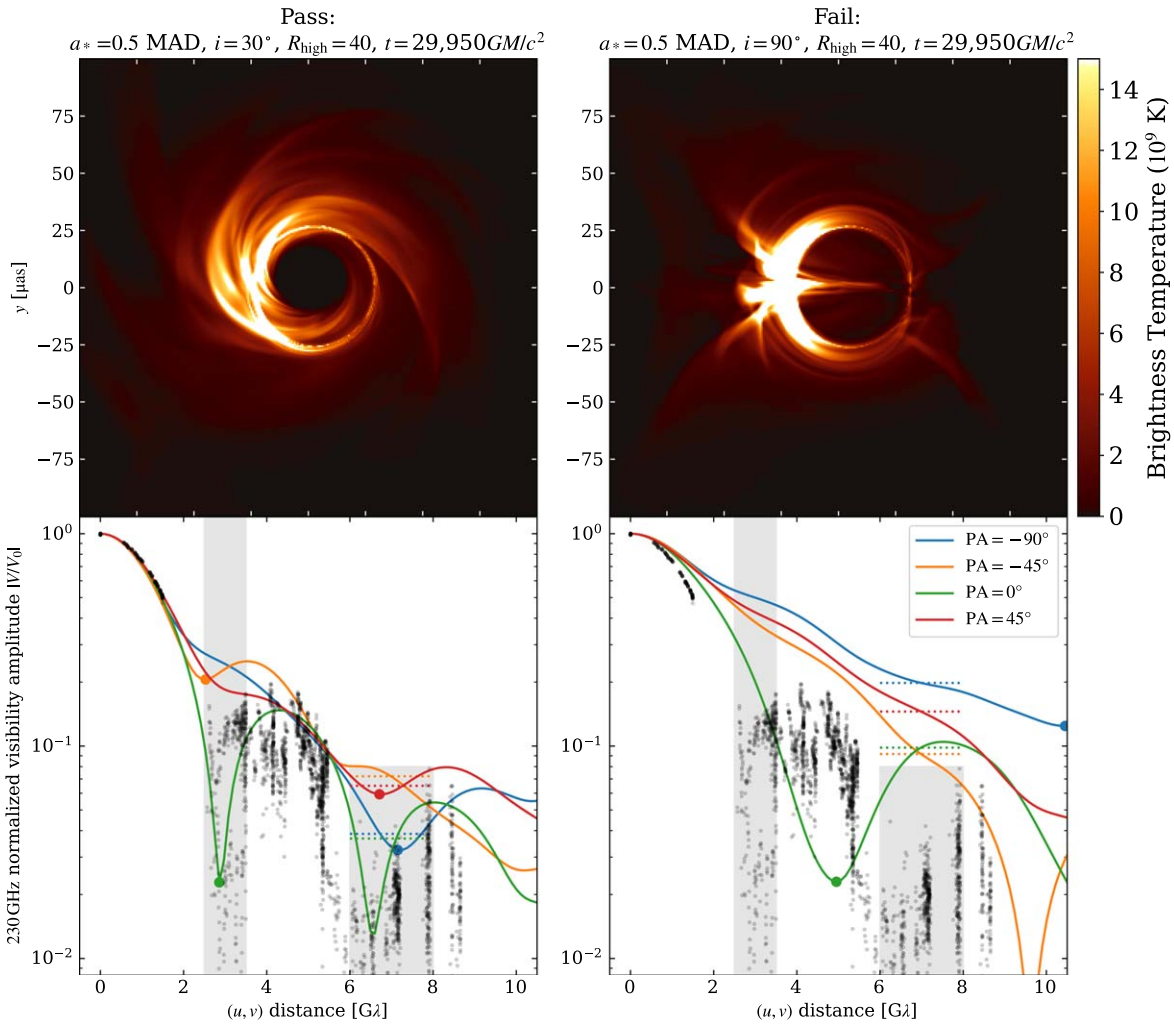

Figure 6. VA morphology constraint example. Top: snapshot images; bottom: VA. Left: passing snapshot; right: failing snapshot. In the bottom row, the solid lines in each plot show VAs on a section through the origin in the (u, v) domain, at four PAs, where 0° is parallel to the projected angular momentum vector of the accretion flow. Filled black points show data from April 7. The VA morphology constraint requires that for at least one PA the first minimum in VA falls within the left gray band, and for all PAs the median of the VAs lies inside the right gray band (see Section 4.1.1, for details). Evidently the snapshot at right fails *both* conditions. The top row shows the corresponding snapshots in the image domain. The color bar on the right is for both images.

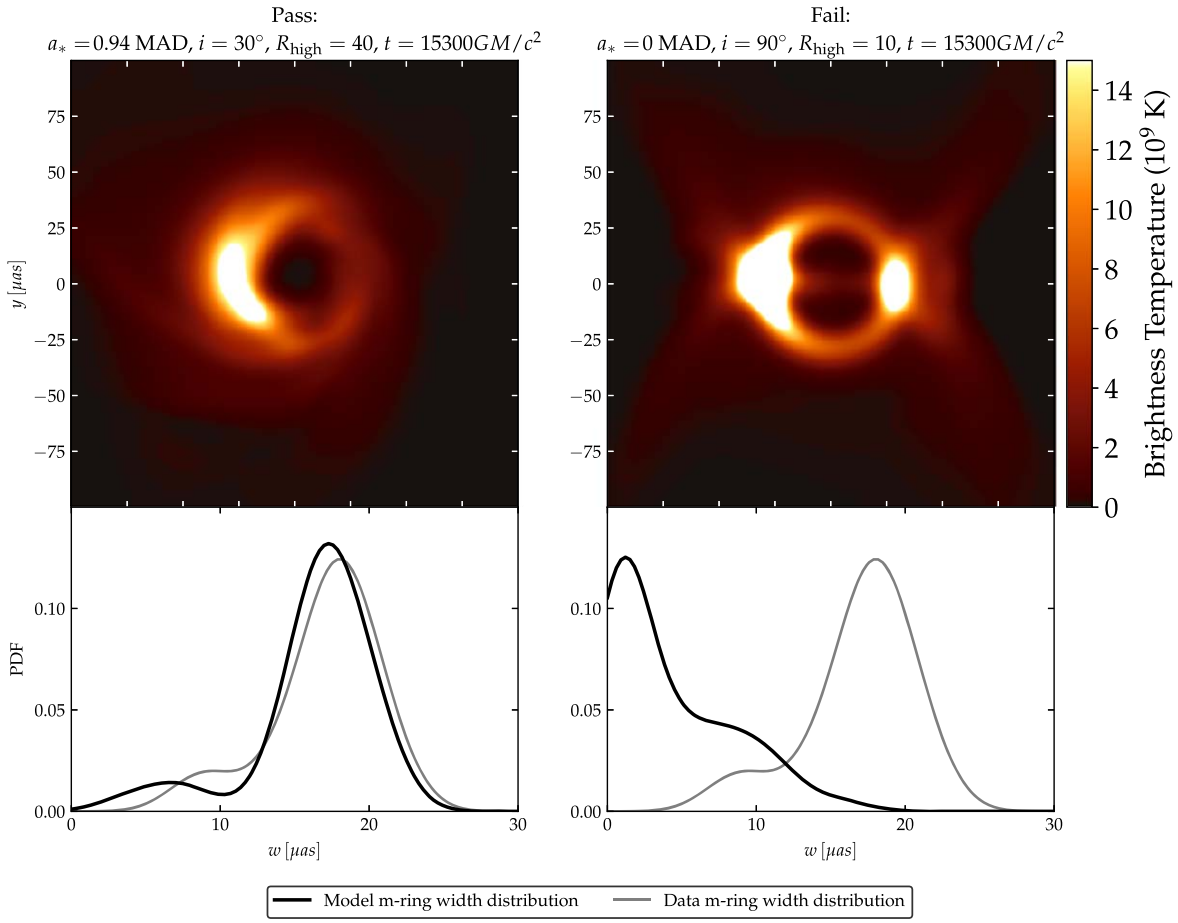

Figure 7. M-ring width constraint example. Top left: snapshot from model that satisfies the constraint, Gaussian blurred to 20 μas; top right: snapshot from model that fails the constraint, Gaussian blurred to 20 μas; bottom left: observed distribution of m-ring widths over our 10 selected scans (light gray) and passing model distribution over 10 scans and four PAs (black); bottom right: observed distribution of m-ring widths over our 10 selected scans (light gray) and failing model distribution (black) over 10 scans and four PAs.

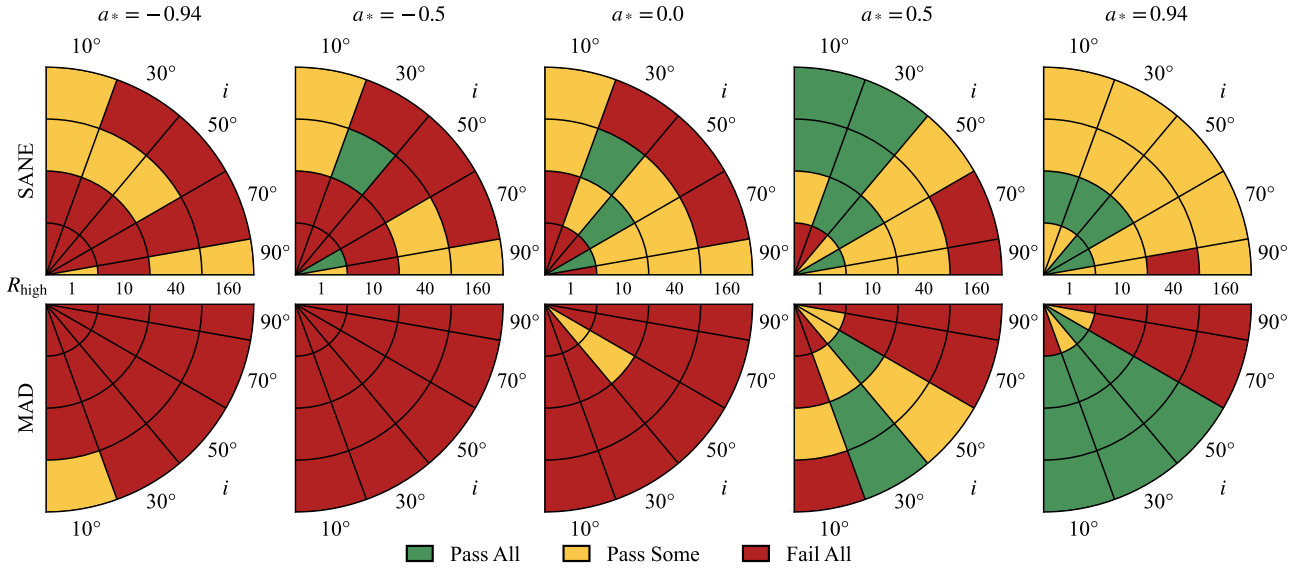

Figure 8. Pass/fail plot for the m-ring widths. Green indicates that the KHARMA, BHAC, and H-AMR models pass, yellow that one or two of the fiducial models fail, and red that all three fail. The inclination coverage is not uniform: BHAC and KHARMA models cover all five inclinations, while H-AMR models cover $i = 10°$, 50°, and 90° only. The $i = 30°$ and 70° wedges therefore include only the BHAC and KHARMA models.

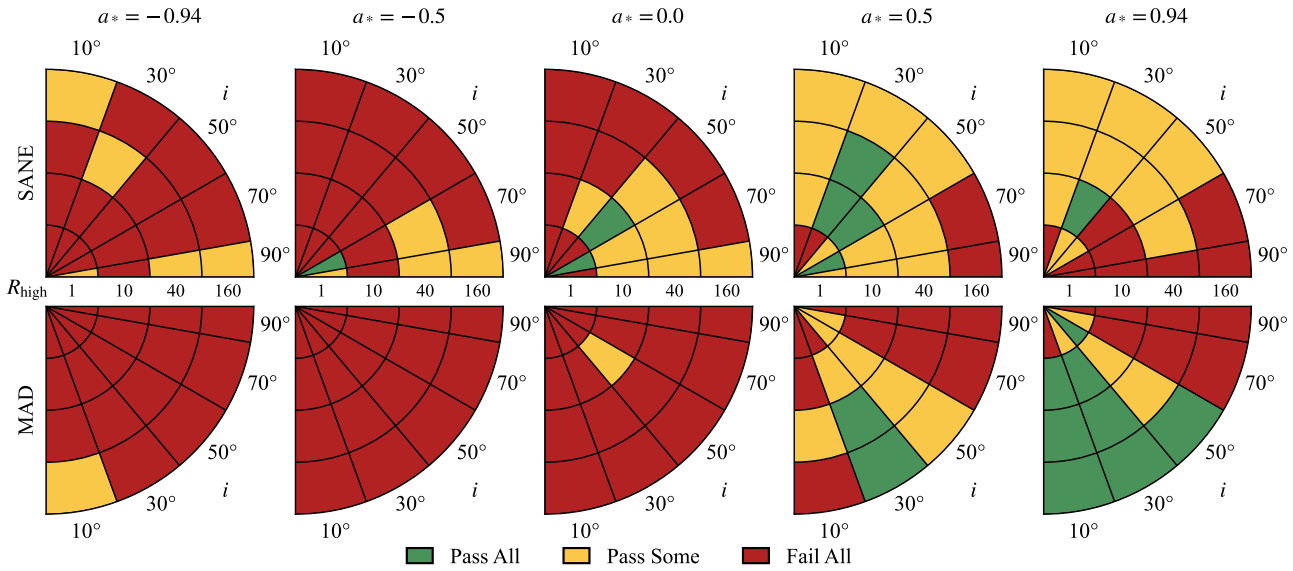

Figure 9. Combined EHT constraints (logical *and*) including the second moment, VA morphology, and m-ring fit constraints. Green indicates that the KHARMA, BHAC, and H-AMR fiducial models pass, yellow that one or two of the fiducial models fail, and red that all three fail. The inclination coverage is not uniform: BHAC and KHARMA models cover all five inclinations, while H-AMR models cover $i = 10°$, 50°, and 90° only.

although ultimately both models are ruled out when extended to 15,000 GM/c^3.

4.2.3. Constant-κ Models with $\kappa = 5$

Next, we consider a model in which all electrons are in a κ eDF, which has a thermal core and a power-law tail. We set $\kappa = 5$ everywhere, motivated by Kunz et al. (2016), who found $\kappa = 5$ to be a good fit to the ion DF in a 3D hybrid simulation of MHD turbulence. A similar application of κ eDFs with fixed κ values for Sgr A* can be found in Davelaar et al. (2018). The power-law tail has $p = \kappa - 1 = 4$, and at high frequency $\nu L_\nu \sim \nu^s$, where $s = 2 - \kappa/2 = -1/2$.[161] We image BHAC GRMHD simulations from 25 to 30 kM using BHOSS (Younsi et al. 2012, 2020). The accretion rate required to obtain 2.4 Jy is smaller than for the thermal models. This implies that many of the $\kappa = 5$ models are optically thin at 230 GHz and show thinner rings than their thermal counterpart (see first and second rows in Figure 14).

230 GHz Size and Light-curve Variability.—We find that the $\kappa = 5$ models produce results that are generally consistent with the BHAC thermal models. Especially at 230 GHz we find similar passing fractions for the 230 GHz source sizes. A total of 92% of the $\kappa = 5$ models pass the size constraint, compared to 98% for the thermal models. This can be explained mainly by SANE models at small $R_{\rm high}$, which are larger than the thermal models. Variability is almost completely unaffected by the κ distribution. We find that 29% are in agreement with the M_3 constraint, compared to 27% for the thermal models. The $\kappa = 5$ models have a higher M_3 for a small number of SANE $R_{\rm high} \geqslant 40$ models. However, since the M_3 constraint is computed for a time window of length only 5000 GM/c^3, a factor of three shorter than for the thermal models, this increase does not increase the fraction of models ruled out by this constraint.

Visibility Amplitude Morphology.—The $\kappa = 5$ models are optically thinner than the corresponding thermal models and typically show a thin, bright ring feature. As a consequence, only 59% of the $\kappa = 5$ models pass the VA morphology constraint, while the passing fraction for the thermal models is 84%. Similarly, due to the change in optical depth, only 55% of the nonthermal models are in agreement with the VA morphology, in contrast to 72% of the thermal models.

M-ring Fits.—The m-ring constraints on diameter, width, and asymmetry are passed by 71%, 3%, and 73% of the $\kappa = 5$ models, respectively. Except for the diameter, all pass fractions are smaller than for the thermal models (65%, 21%, and 95% for diameter, width, and asymmetry, respectively). The slightly larger pass fraction for the diameter could be affected by the shorter time window used for the κ models as compared to the thermal ones. However, the low fraction for the m-ring width can be explained by the optical depth of the κ models. Most of the κ models are optically thinner than their thermal counterpart, which leads to a finer, brighter ring structure, and this is picked up by the m-ring fitting (see Figure 14).

86 GHz Source Size.—For MAD models the size of the $\kappa = 5$ models does not change. This can be explained by the fact that most of the emission is produced in the midplane. For the SANE models we find two different behaviors: the source size increases for $R_{\rm high} < 40$ and decreases for $R_{\rm high} \geqslant 40$, especially for positive black hole spins and high inclinations (compare first and last panel in the bottom row of Figure 14). This change in size is consistent between the images at 230 GHz and at 86 GHz. The passing fraction for the $\kappa = 5$ models drops to 29% as compared to 59% for the thermal models.

86 GHz Flux.—The $\kappa = 5$ models are relatively optically thin at 86 GHz. Together with the spectral slope $p = \kappa - 1$, the flux at lower frequencies can be approximated as $2.4 \times (\nu/230\ {\rm GHz})^{-(p-1)/2}$. This leads to an 86 GHz flux density ∼10 Jy, which is far above the 86 GHz flux constraint of 2 ± 0.2 Jy. Consequently, the passing fraction for the $\kappa = 5$ models drops to 12%, compared to 68% for the thermal models.

[161] Unless stated otherwise, the width parameter w of the κ distribution (see Equation (18)) is set by $w = (\kappa - 3)\Theta_e/\kappa$, where the dimensionless electron temperature Θ_e is computed according to Equation (13).

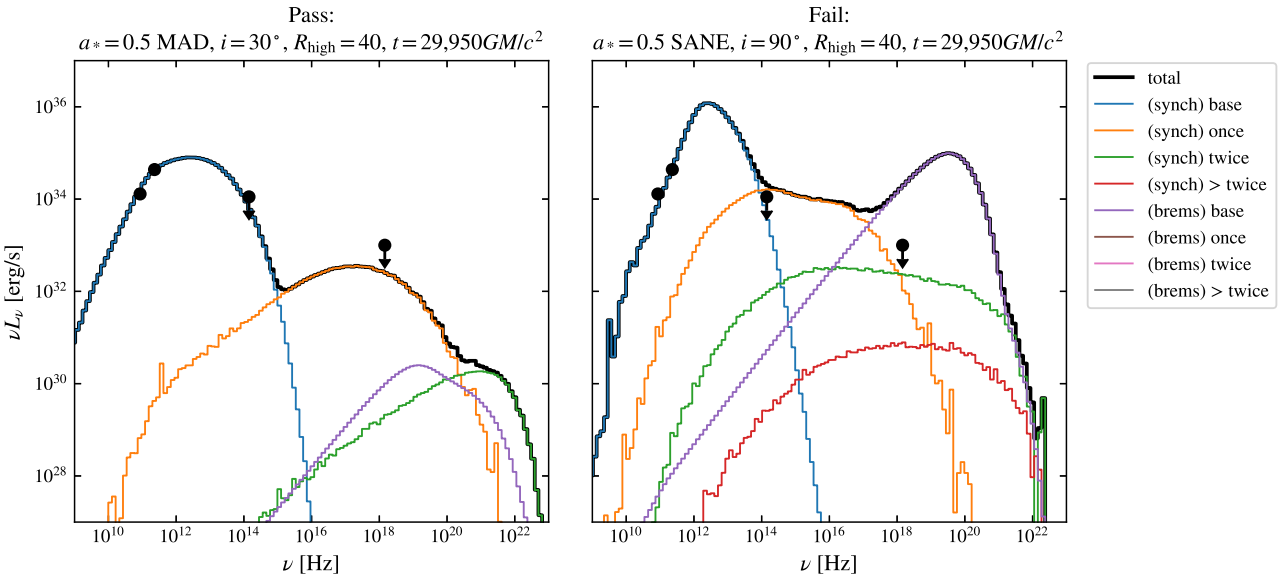

Figure 10. Non-EHT flux density constraint example. Left: passing model with SED close to the measured 86 GHz point and below the quiescent 2.2 μm and X-ray points. Right: failing model with inconsistent (strongly rising) millimeter wavelength spectral index, overproduction of 2.2 μm due to strong Comptonization, and overproduction of X-rays by bremsstrahlung.

2.2 μm Constraint.—All MAD models fail the 2.2 μm constraint, as do SANE models with $R_{\rm high} > 1$. This can be explained by the power-law tail of the κ eDF (see Equation (15)) as compared to the exponential behavior in the thermal eDF (see Equation (11)). Only 14% of the κ models pass, in contrast to 70% of the thermal models. Evidently the NIR flux density provides a powerful constraint on any nonthermal component in the eDF.

4.2.4. Mixed Thermal/κ Model

Next, we consider a mixed thermal/nonthermal eDF, with the nonthermal component following the κ DF with $\kappa = 3.5$. At high frequency $\nu L_\nu \sim \nu^s$ with $s = 2 - \kappa/2 = 1/4$, similar to what is seen in 2.2 μm flares (Hornstein et al. 2007). For this model set the GRMHD simulations use BHAC and the imaging uses BHOSS.

The fraction of nonthermal electrons is assumed to depend on σ and β. The emissivity

$$j_{\nu,\rm tot} = (1 - \epsilon) j_{\nu,\rm thermal} + \epsilon j_{\nu,\kappa}, \tag{16}$$

where the nonthermal efficiency

$$\epsilon(\varepsilon, \beta, \sigma) = \varepsilon[1 - e^{-\beta^{-2}}][1 - e^{-(\sigma/\sigma_{\rm min})^2}]. \tag{17}$$

Evidently $\epsilon \to 0$ in the disk while $\epsilon \to \varepsilon$ in the jet. Since we remove emission at $\sigma > \sigma_{\rm cut} = 1$, the nonthermal electrons are confined to the jet sheath.

We set $\sigma_{\rm min} = 0.01$ and vary the base efficiency, ε, over 0.05, 0.1, and 0.2. At each ε we generate a model set spanning the same parameter space as the thermal models (see Table 2) and normalize the accretion rate using the standard procedure (see Section 2).

The mass accretion rate required to obtain 2.4 Jy at 230 GHz only changes on average around 1.5% as compared to the thermal models. This small variation in the mass accretion rate reveals the fact that most of the emission at 230 GHz is created from the thermal part of the hybrid eDF, consistent with the small fraction of nonthermal particles added ($\varepsilon = 0.05$, 0.1, and 0.2).

230 GHz Size and Light-curve Variability.—The addition of nonthermal particles does not substantially affect the flux or size of the image at 230 GHz.

For MAD models, 230 GHz emission is mostly produced in the disk region (see M87* Paper V and Figure 8 in Wong et al. 2022, for a 3D rendering). Thus, the images are unaffected by the nonthermal particles, which are located in the jet.

For SANE models, increasing $R_{\rm high}$ pushes the emission toward the jet sheath, which increases the source size for high spins and large $R_{\rm high}$. However, the effect of nonthermal particles on the image is minor because most of the emission is still produced by thermal electrons with temperature set by the $R_{\rm high}$ prescription (compare first and third panels in the top row of Figure 14).

The passing fraction for the 230 GHz image size is 98%, independent of ε, consistent with the thermal models. We find that 47% of the models are in agreement with the 230 GHz variability constraint. This passing fraction is larger than for the thermal models (27%), due to the shorter time window ($5000\,GM/c^3$) considered for the nonthermal models, in contrast to $15{,}000\,GM/c^3$ for the thermal ones.

Visibility Amplitude Morphology and Variability.—Since 230 GHz images of the $\kappa = 3.5$ models with variable efficiency are similar to the thermal models (see previous paragraph), the fraction of passing models for the VA morphology are comparable. The three nonthermal models have an average passing fraction for the VA morphology of 80%, whereas 84% of the thermal models pass. The 4 Gλ VA variability constraint passes 72% of the models for both thermal and nonthermal eDF.

M-ring Fits.—Given that including nonthermal particles via the equations presented in Equation (17) does not change the image structure and variability properties of the 230 GHz images, the M-ring fits provide the same passing fractions for the diameter (65%), width (22%), and asymmetry (95%).

86 GHz Source Size and Flux.—The 86 GHz source is only slightly affected by the addition of nonthermal particles as compared to the thermal models. Only the SANE models with $R_{\rm high} \geqslant 40$ and $a_* > 0$ produce 86 GHz image sizes larger than

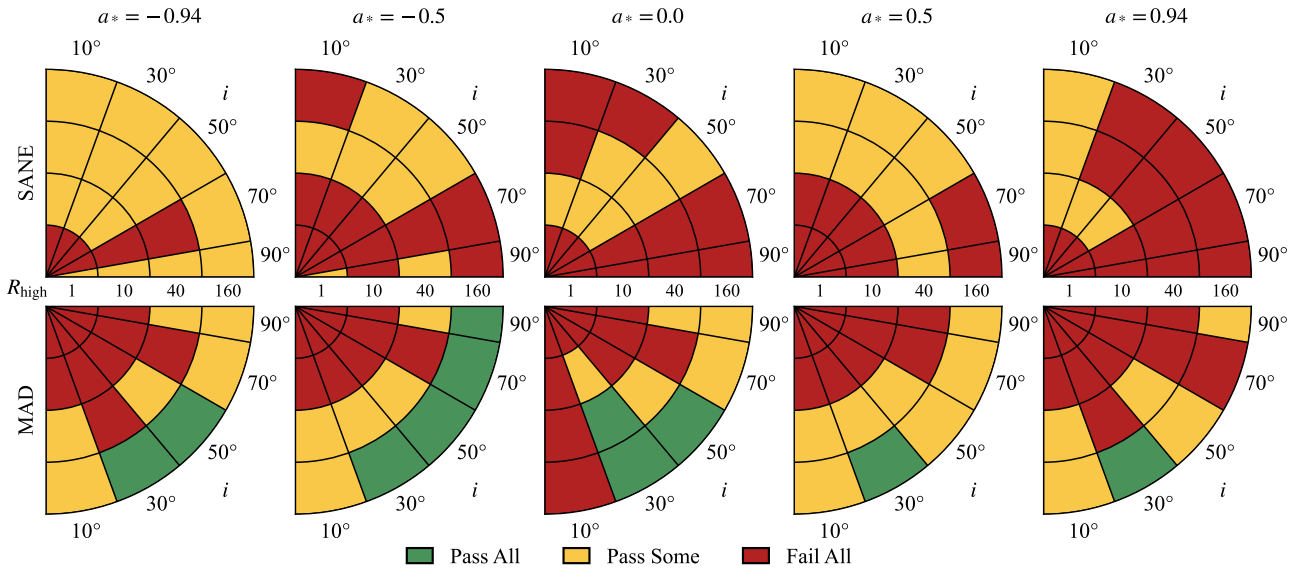

Figure 11. Combined non-EHT constraints (logical *and*). Green indicates that the KHARMA, BHAC, and H-AMR models pass, yellow that one or two of the fiducial models fail, and red that all three fail.

the thermal SANE models. This effect can be seen in the first and third panels in the bottom row of Figure 14. Notice the increased flux density in the jet sheath in the difference image (blue color). This trend increases with the efficiency and is reflected in the decreasing pass fraction: 56% (for $\varepsilon = 0.05$, 0.1) and 55% ($\varepsilon = 0.2$) as compared to thermal models (59%). A similar trend is found for the 86 GHz flux density. The nonthermal particles are mainly located in the jet and thus contribute to the 86 GHz flux. Again, jet-dominated high-spin SANE models typically fail the 86 GHz flux constraint. With increasing efficiency, i.e., adding more nonthermal particles, the pass fraction decreases, with 67% passing at $\varepsilon = 0.05$, 66% passing at $\varepsilon = 0.1$, and 63% passing at $\varepsilon = 0.2$, compared to a pass fraction of 68% for the thermal models.

2.2 μm Constraint.—The 2.2 μm flux density increases for all models. For SANE models, except $R_{\rm high} = 1$, the addition of nonthermal particles leads to overproduction of 2.2 μm photons. For MAD models, all models overproduce at 2.2 μm for $\varepsilon \geqslant 0.05$. As noted above, 2.2 μm emission is produced from the tail of the eDF. The thermal eDF tail decreases exponentially, while the κ eDF tail decreases as a power law, so the increase in 2.2 μm flux density is unsurprising. This is a general feature of the nonthermal models: 2.2 μm observations sharply limit the allowed population of nonthermal electrons.

4.2.5. Variable-κ Model

The high-energy variability observed in many astrophysical sources, including the Galactic center, may be associated with magnetic reconnection. Particle-in-cell (PIC) simulations have found that the slope of the nonthermal tail depends on σ and β (see, e.g., Ball et al. 2018). Here we consider a κ eDF model in which κ and w vary following the prescription of Ball et al. (2018):

$$\kappa = 2.8 + 0.7\sigma^{-1/2} + 3.7\sigma^{-0.19}\tanh(23.4\,\sigma^{0.26}\beta), \quad (18)$$

$$w = \frac{\kappa - 3}{\kappa}\Theta_e. \quad (19)$$

We use emissivities and absorptivities from Pandya et al. (2016), computed numerically for the interval $3 < \kappa \leqslant 8$. For $\kappa > 8$ we substitute a thermal eDF. As in the fiducial models, we turn off emission at $\sigma > 1$.

The variable-κ models are computed from H-AMR and BHAC GRMHD models, where the time windows 30,000–35,000 GM/c^3 (H-AMR) and 25,000–30,000 GM/c^3 (BHAC) are used.

We find that the mass accretion rate needed to obtain $\langle F_{230}\rangle = 230$ GHz is on average 4% larger than for the thermal models, and thus the variable-κ models have slightly higher optical depth.

230 GHz Size.—The disk region is dominated by thermal electrons (i.e., large κ), while the jet sheath has the lowest κ. Therefore, the 230 GHz source size of the variable-κ models is similar to the thermal ones, and no difference in pass fraction is found. A total of 98% of both models are in agreement with the 230 GHz size estimate (see the first and second panels in the top row of Figure 14).

Visibility Amplitude Morphology and Variability.—For the null location of the variable-κ models we find no difference to their thermal counterparts, and for both ~80% pass this constraint. However, there is a clear discrepancy between the variable-κ and thermal models regarding the VA morphology. Only 60% of the κ models pass the 4 Gλ VA variability constraint, in contrast to 72% of the thermal models.

M-ring Fits.—The thermal and variable-κ models agree in the passing fraction for the m-ring diameter (66%), m-ring width (22%), and asymmetry (95%).

86 GHz Source Size and Flux.—The pass fractions for the 86 GHz source size constraint are comparable for thermal (59%) and variable-κ models (55%). Given that most of the variable-κ models are optically thicker than their thermal counterparts, the 86 GHz flux is on average lower, which increases the passing fraction from 68% (thermal eDF) to 75% (variable κ eDF).