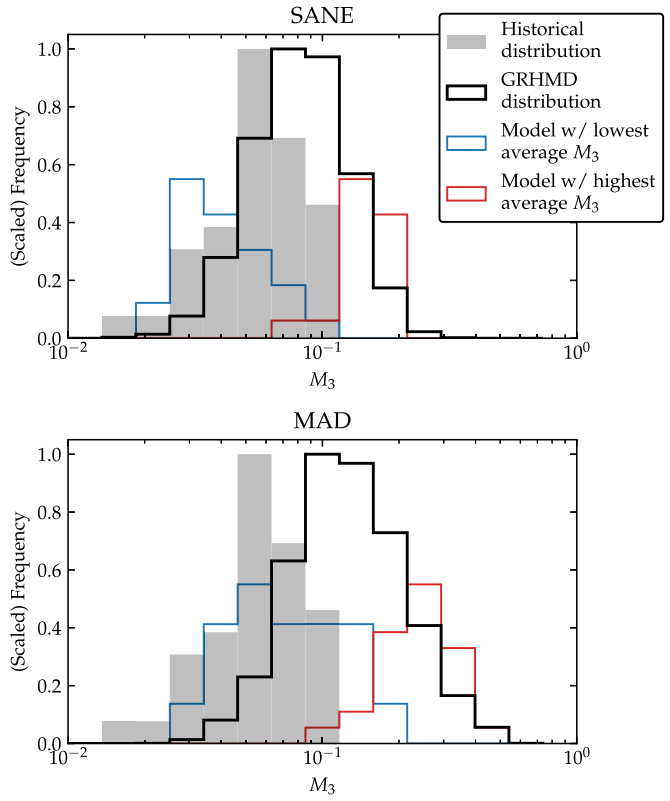

Figure 12. Distributions of 3 hr modulate index M_3 for BHAC, KHARMA, and H-AMR models (black), compared to distributions from historical observations (gray). The distributions for models with the lowest (blue) and highest (red) average M_3 for SANEs and MADs are also shown. The heights of these distributions have been scaled for clarity.

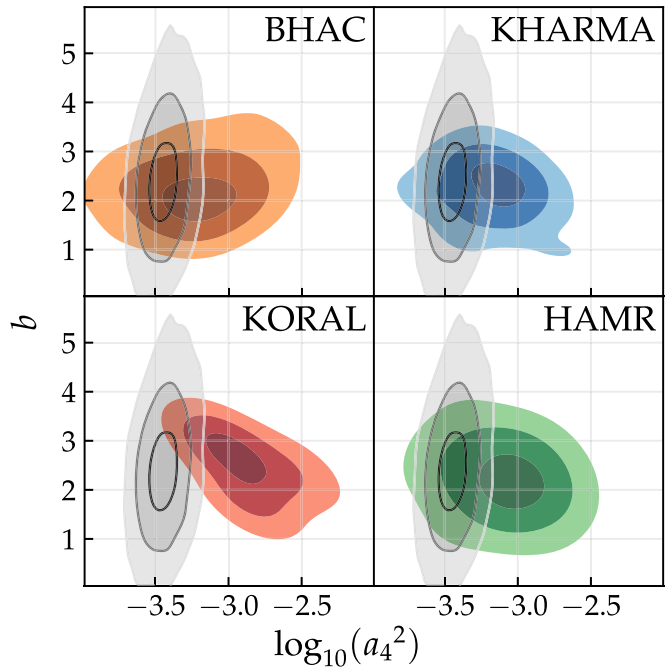

Figure 13. Distributions of ($\log_{10}(a_4^2)$, b) for BHAC, KHARMA, koral, and H-AMR, compared against the 2017 EHT confidence regions (gray).

2.2 μm Constraint.—Including nonthermal particles via the κ eDF increases the 2.2 μm flux as compared to the thermal eDF. In our prescription κ decreases (more high-energy electrons) as σ increases. In MAD models σ is systematically larger than in SANE models. Consistent with this, we find that most of the MAD models fail the 2.2 μm constraint, whereas in SANE models κ is large and the passing fraction for SANE models is almost indistinguishable from their thermal counterparts. In total, 35% of the variable-κ models pass the 2.2 μm constraint, compared to 55% for models including only thermal particles.

X-Ray Constraint.—Including nonthermal particles via the variable κ eDF reduces the pass fraction from 61% (for thermal eDF) to 35%. The κ eDF provides a larger population of seed photons in the NIR and at higher energies that can be boosted into the X-ray by a single scattering event.

Trends across BHAC *and* H-AMR *Models for Variable-*κ *Models.*—For the variable-κ models we have redundant models from BHAC and H-AMR (see Table 1). Both model sets show similar trends for all constraints (see Table 6).

4.2.6. Summary of Constraints on Nonthermal Models

In Table 6 we list the pass fractions for BHAC and H-AMR models using different eDFs. Most nonthermal eDF models produce little change compared to the thermal models for most constraints. The 86 GHz size and flux, which are the most important non-EHT constraints, are only marginally affected by the addition of nonthermal electrons. This behavior is obtained especially for eDFs, which mainly add nonthermal particles in the jet, while the disk is populated by thermal ones. In our case this setup is given for variable κ and $\kappa = 3.5$ with variable efficiency and is consistent between BHAC and H-AMR models (see Table 6).

If nonthermal particles are included also in the disk, either via a power law with slope $p = 4$ stitched to a thermal distribution or via a $\kappa = 5$ distribution, then there are some variations in pass fractions as compared to the above-mentioned eDFs. For the power-law models the addition of nonthermal electrons increases the 86 GHz size by an average of 50%. However, the pass fractions with respect to the thermal models are not changed. In contrast, the $\kappa = 5$ model pass fractions decrease by 20% compared to the thermal models. For $p = 4$ and $\kappa = 5$, fewer models pass the 230 GHz m-ring width, with a consistent decrease by ~20% for both models. Interestingly, the other m-ring constraints, i.e., the diameter and asymmetry, are not affected by the addition of nonthermal particles. This can be explained by the finer and brighter ring feature found in $p = 4$ and $\kappa = 5$ models connected to their smaller optical depth compared to their thermal counterparts (see Table 6).

In general, the fraction of models passing M_3 increases with the addition of nonthermal particles, independent of the prescriptions of the eDF. This is due to the shorter duration of the exploratory runs and not an actual reduction in variability. The main characteristic of the nonthermal models is the increase of 2.2 μm and X-ray flux densities. However, in a large fraction of models this leads to overproduction of 2.2 μm or X-ray flux and the pass fractions are reduced.

Six nonthermal models pass all 11 constraints (see Table 7). These models are one H-AMR MAD $p = 4$ model with $a_* = 0.5$ seen under an inclination $i = 50°$ and $R_{\rm high} = 1$ and a high spinning SANE model with $a_* = 0.94$ at an inclination of $i = 50°$ with $R_{\rm high} = 40$. From the BHAC variable κ three models are in agreement with all constraints, namely, spin $a_* = 0.5$ at inclination $i = 10°$ at $R_{\rm high} = 80$ and 160 and a model with the same spin seen under a slightly larger angle of

Table 5
Fiducial Models That Fail Only One Constraint

Code/Setup	MAD/SANE	Spin a_*	Inclination i	$R_{\rm high}$	Failed Constraint
KHARMA thermal	SANE	0.94	10	40	86 GHz size
KHARMA thermal	SANE	0.94	30	40	86 GHz size
KHARMA thermal	SANE	0.94	50	1	86 GHz size
KHARMA thermal	MAD	0.5	30	40	M_3
KHARMA thermal	MAD	0.5	30	160	M_3
KHARMA thermal	MAD	0.94	10	160	M_3
KHARMA thermal	MAD	0.94	30	160	M_3
BHAC thermal	SANE	−0.5	30	40	M-ring diameter
BHAC thermal	SANE	0	30	40	M-ring diameter
BHAC thermal	SANE	0.5	10	40	M_3
BHAC thermal	SANE	0.5	10	160	M_3
BHAC thermal	SANE	0.5	30	40	M_3
BHAC thermal	SANE	0.5	30	160	M_3
BHAC thermal	MAD	0.5	30	160	M_3
BHAC thermal	MAD	0.5	50	160	M_3
BHAC thermal	MAD	0.94	10	160	M_3
BHAC thermal	MAD	0.94	30	160	M_3

Note. Models that pass all but one constraint. Since no model passes all constraints, these represent the parameters that are closest to being consistent with observations.

$i = 30°$ with $R_{\rm high} = 160$. The last of the six survivors is a BHAC SANE model with variable efficiency of $\varepsilon = 0.05$ with an inclination of 10° and $R_{\rm high} = 10$. These models share a common low inclination angle $i \leqslant 60°$ and positive spin. We note that the MAD models coincide with the cluster of thermal models found for both BHAC and KHARMA models (see Section 4.1.4).We also show the nonthermal models that fail only one constraint in Table 8.

4.3. Tilted Models

Aligned models are a special case: in general, one expects that the spin angular momentum of the black hole and the orbital angular momentum of the accretion flow are misaligned. Here we consider misaligned flows around an $a_* = 15/16$ black hole from Liska et al. (2018) and Chatterjee et al. (2020).

All aligned models considered so far produce either a SANE or MAD accretion flow. The tilted disk model initial conditions, however, produce a strongly magnetized near-MAD outcome with dimensionless magnetic fluxes between 25 and 50, a state we describe as IN-SANE. We consider three GRMHD simulations with tilt 0°, 30°, and 60°.

The tilted models exhibit a warped disk due to Lense–Thirring precession. The time-averaged disk and jet are therefore nonaxisymmetric. Since the inner and outer disks have different orientations, it is necessary to specify the coordinate axis of the observer. We consider three observer inclinations with respect to the *outer* disk at a single azimuth of 0° (for more details, see Chatterjee et al. 2020).[162]

The 230 GHz pre-image size of edge-on large-$R_{\rm high}$ models increases slightly for the tilt-60° compared to the aligned case. This occurs because the inner jet is warped and creates an extended image. This effect is also seen in the 86 GHz image size. On the other hand, the 86 GHz flux varies little with tilt despite the presence of a boosted jet component at large tilt angles.

[162] A full parameter survey would run over azimuth angle as well.

Variability increases with tilt. In tilted disks, accretion occurs via thin plunging streams (e.g., Fragile et al. 2007), where electrons in the shocked flow can be heated to relativistic temperatures (e.g., Dexter & Fragile 2013; Generozov et al. 2014; White et al. 2019), forming localized, fluctuating hot spots more easily than in aligned disks and increasing flux variability (Chatterjee et al. 2020; Bauer et al. 2022). Nevertheless, 20/27 models pass the M_3 constraint because of the short duration of the tilted models, which provide fewer M_3 samples than the fiducial models.

The 2.2 μm flux density also increases with tilt. The 2.2 μm flux exceeds the 1.0 mJy limit for all three tilts, with a few exceptions, e.g., $R_{\rm high} = 160$ models at 10° inclination, which makes it difficult to favor the aligned case over the tilted one. Furthermore, misalignment destroys the axisymmetric nature of the accretion flow. The current model set covers a small parameter space in inclination and $R_{\rm high}$. A thorough exploration of the source azimuthal angle with respect to the observer is left to future studies.

To summarize: for the model set considered here tilt primarily affects variability and the 2.2 μm flux density, tending to increase both and thus shifting acceptable aligned models into rejected models as tilt angle increases. These trends are consistent with those observed by White & Quataert (2022).

4.4. Stellar-wind-fed Models

The accretion models of Ressler et al. (2020a, 2020b, 2018) track plasma from magnetized stellar winds down to the event horizon and provide a self-consistent picture of the origin of both gas and magnetic fields in the accreting plasma in Sgr A*. The resulting inflow does not fully circularize, so the models provide a distinct alternative to the fiducial models, which *assume* that the torus initial conditions relax to an astrophysically accessible state for the inner accretion flow. In the wind-fed models the density of the wind is fixed, so the 230 GHz flux density is matched to observations by varying $R_{\rm high}$ instead.

Table 6
Exploratory Model Pass Fractions

Constraint/Model	BHAC				H-AMR		
	Thermal	$\kappa(\sigma, \beta)$	$\kappa = 3.5$ $\varepsilon = 0.05, 0.1, 0.2$	$\kappa = 5$	Thermal	$\kappa(\sigma, \beta)$	$p = 4$
230 GHz size	0.98	0.99	0.98, 0.98, 0.98	0.92	1.0	0.99	0.94
VA morphology	0.83	0.80	0.81, 0.81, 0.78	0.59	0.80	0.88	⋯
M-ring diameter	0.65	0.69	0.66, 0.66, 0.67	0.71	0.58	0.67	0.89
M-ring width	0.21	0.21	0.24, 0.23, 0.23	0.03	0.29	0.40	0.13
M-ring asym.	0.95	0.97	0.95, 0.95, 0.94	0.73	1.0	0.97	0.98
86 GHz flux	0.68	0.75	0.67, 0.66, 0.63	0.12	0.62	0.65	0.72
86 GHz size	0.59	0.57	0.56, 0.56, 0.55	0.38	0.46	0.45	0.47
2.2 μm flux	0.55	0.35	0.14, 0.12, 0.12	0.14	0.80	0.2	0.41
X-ray flux	0.70	⋯	⋯	⋯	0.61	0.35	⋯
Light-curve variability	0.27	0.30	0.47, 0.47, 0.46	0.29	0.07	0.28	0.41
4 Gλ variability	0.72	0.60	0.74, 0.73, 0.71	0.55	0.39	0.46	0.63
EHT constraints	0.19	0.17	0.17, 0.16, 0.15	0.01	0.22	0.27	0.10
Non-EHT constraints	0.19	0.12	0.01, 0.0, 0.0	0.14	0.22	0.09	0.39
Variability constraints	0.27	0.28	0.42, 0.42, 0.42	0.27	0.03	0.24	0.38

Note. Pass fractions for BHAC and H-AMR models for various eDFs. Note that the thermal models are run for 10,000 GM/c^3 and 15,000 GM/c^3 for BHAC and H-AMR, respectively, while the nonthermal models are only run for 5000 GM/c^3. This results in a much lower pass fraction in the thermal models for the light-curve variability constraint and all three m-ring constraints.

We use two versions of the model: one in which the stellar wind magnetization is low ($\beta = 10^6$), and a second in which the magnetization is high ($\beta = 10^2$). $R_{\rm high}$ is adjusted until each model has the observed time-averaged 230 GHz flux density, with $R_{\rm high} = 13$ ($\beta = 10^6$) and $R_{\rm high} = 28$ ($\beta = 10^2$).

Both wind-fed models produce rings that are too narrow, failing the m-ring width test. In addition, both are too bright at 86 GHz and fail the M_3 test, although they are quieter than MAD models and close to the cutoff.

Both non-EHT and EHT constraints have the power to test wind-fed models. It is *not* possible to draw broad conclusions about the viability of the wind-fed models in general, however, since the two models tested here contain only a single spin ($a_* = 0$) and all use the $R_{\rm high}$ thermal eDF model.

5. Discussion

5.1. The Goldilocks Model and Polarization

The fiducial models cover a regular grid in parameter space. In one instance, for KHARMA models, adjacent points in parameter space fail only one constraint for opposite reasons: the SANE, $a_* = 0.5$, $R_{\rm high} = 40$, $i = 10°$ models fail because the 86 GHz image is too small, while the $i = 30°$ models fail because the 86 GHz image is too large, suggesting that a model with intermediate inclination would pass all constraints.

We analyzed an intermediate-inclination model at $i = 20°$. This "goldilocks" model passes *all* constraints (we did not compute the X-ray luminosity, but the neighboring $i = 10°$ and $i = 30°$ models pass the X-ray constraint).

We have imaged a series of KHARMA models with inclinations between 10° and 30°. The cause of the inclination sensitivity is an extended, 86 GHz bright jet. At $i = 10°$ the jet is nearly parallel to the line of sight and the source size is dominated by the accretion flow, but as i increases, the jet, which is radially extended, begins to extend past the accretion flow and dominate the source size.

Despite this success, we regard the goldilocks model as unpromising for two reasons. First, the 10° and 30° BHAC models fail the X-ray constraint, and the neighboring 10° and 50° H-AMR models fail several constraints.

In addition, the goldilocks model is likely underpolarized at 230 GHz. The source-integrated linear polarization of Sgr A* at the time of the 2017 campaign was between 6.9% and 7.5% (Goddi et al. 2021), consistent with historical measurements. We will consider linear and circular polarizations in future papers, but our preliminary finding is that the goldilocks model has linear polarization ≈1%, which is far too low. We find that the best-bet region MAD models, considered below, have linear polarization compatible with observations.

5.2. Testing Exploratory Models That Pass

Six exploratory models, considered in Section 4.2, pass all constraints. Although they appear promising, these six models are imaged for only 5000 GM/c^3 and thus have been tested more weakly than the fiducial thermal models.

To evaluate the effect of run duration, we imaged the six passing exploratory models for 15,000 GM/c^3. All failed one or more constraints, which are listed in Table 7. Evidently adding a population of nonthermal electrons does not provide a consistent way of transforming failing fiducial models to passing models.

The pass/fail status of the model can be sensitive to the length of integration, and it is important to image the models for at least 15,000 GM/c^3. In connection with this, we note that the koral models, which were imaged at 230 GHz for $\sim 10^5 GM/c^3$, were generally consistent with the fiducial models but provide tighter M_3 constraints (see Appendix B). The constraints that are most sensitive to model duration are M_3 (all models failed M_3 after being extended) and the m-ring fits (two failed m-ring width, one failed m-ring diameter).

5.3. Origin of Variability Excess

Approximately 84% (KHARMA), 73% (BHAC), and 97% (H-AMR) of the fiducial models fail one or both variability

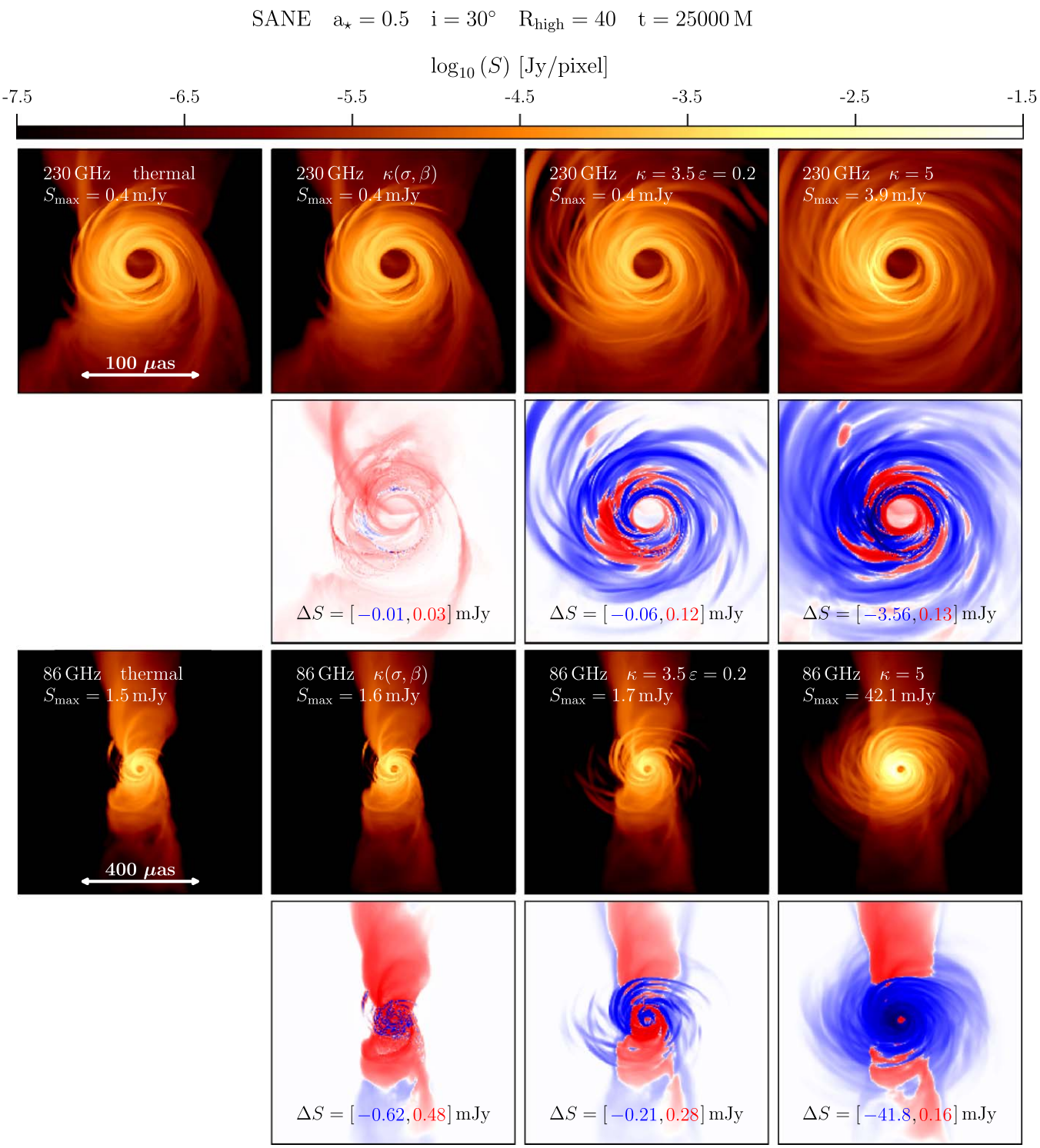

Figure 14. Influence of the eDF on the image structure for a SANE model with spin $a_\star = 0.5$ seen under a viewing angle of 30° using $R_{high} = 40$ at $t = 25{,}000\ GM/c^3$. In the first and third rows the panels show the image structure, from left to right, of a thermal, variable kappa $\kappa(\sigma, \beta)$, $\kappa = 3.5$ (fixed) with efficiency $\varepsilon = 0.2$, and $\kappa = 5$ (fixed) everywhere eDF at 230 GHz and at 86 GHz. Notice the increased field of view for the 86 GHz images. The second and fourth rows show the difference between thermal image and the different nonthermal eDFs.

constraints. This naturally leads one to ask whether there is an observational or modeling problem with these constraints.

For example, it is possible that a fraction f_{ext} of the 230 GHz flux density is in an extended structure (e.g., a jet) that is slowly varying, unresolved by ALMA, and resolved out by EHT. The observed M_3 would then be smaller than the true M_3 for the compact source by a factor of $1/(1+f_{ext})$. The 4 Gλ amplitude variability is normalized by the light curve, and thus a_4 would be suppressed by a similar factor.

Diffuse emission on scales larger than the VLBI images ($\sim 100\ \mu$as) and smaller than connected element interferometer measurements (~ 100 mas) is difficult to constrain. The EHTC

imaging strategy involves a self-calibration step that assumes no diffuse structure on scales between the zero baseline and the shortest VLBI baselines (Paper III). However, longer-wavelength VLBI observations place constraints on any emission on these scales under the assumption of flat or steep spectra. For instance, 230 and 43 GHz VLBI observations that use the Los Alamos-Pie Town-VLA baselines probe scales of ∼1 to 10 mas and demonstrate no inconsistency in closure amplitudes and closure phases with a symmetric, two-dimensional Gaussian model (Bower et al. 2004). VLBI observations at 3 mm wavelength are also fully consistent with a two-dimensional Gaussian model with an upper limit of ∼ 10 mJy, or approximately 1% of the total flux density (Brinkerink et al. 2019). Additionally, a dust contribution is constrained by shorter-wavelength ALMA observations that find a flat or slightly falling spectrum up to 900 GHz (Bower et al. 2019). A substantial diffuse dust contribution would only be consistent if its properties were tuned to match a steeply falling compact synchrotron spectrum, which would likely be inconsistent with the substantially variable far-infrared component of Sgr A* emission (Stone et al. 2016; von Fellenberg et al. 2018).

Future observations with short-baseline coverage such as that provided by the Kitt Peak−to−SMT baseline will enable tighter constraints on any diffuse component. If confirmed, the presence of diffuse flux would require renormalization of the models and reevaluation of the constraints. A reduction of ∼30% in the compact flux would lead many SANE models to fail the M_3 constraint because they would be not variable enough and would lead most MAD models to pass the M_3 constraint. Since $F_{230} \sim \dot{M}^2$, a 15% change in model density normalization, and therefore in optical depth, would suffice.

The variability excess might also be caused by physical incompleteness of the models. Collisionless effects and radiative cooling could both reduce variability.

We model the accreting plasma as an ideal fluid when in fact it is collisionless, with Coulomb mean free path large compared to GM/c^2. This is less worrisome than one might think: electrons and ions are confined to helical orbits around field lines, implying an effective mean free path perpendicular to the field lines of order the gyroradius $\sim 56\Theta_e[B/(30\,\mathrm{G})]^{-1}\,\mathrm{cm} \ll GM/c^2$. The mean free path parallel to field lines is still long but may be limited by scattering off electromagnetic field fluctuations excited by kinetic instabilities rather than Coulomb scattering.

Future global kinetic general relativistic PIC simulations may be able to test how well the ideal fluid model describes collisionless accretion flows. Meanwhile, a relativistic fluid model incorporating small collisionless corrections was developed by Chandra et al. (2015) and studied numerically by Foucart et al. (2017). The leading-order corrections are conduction and pressure anisotropy (i.e., viscosity). The effect of conduction and viscosity on our constraints is not known, but it is known that viscosity reduces turbulent stress in SANE models (Foucart et al. 2017), consistent with a reduction in variability. It is also plausible that conduction smooths out temperature maxima, possibly also leading to a reduction in variability.

Our models neglect radiative cooling. Cooling is fastest where the electron temperature is highest, so cooling has the potential to blunt local maxima in temperature (Yoon et al. 2020). If local maxima with short cooling times contribute significantly to variability, then cooling could reduce M_3. Self-consistent cooling requires integration of an electron energy equation (e.g., Ressler et al. 2015) and assignment of a density scale (mass unit or accretion rate) when the GRMHD simulation is run, vastly increasing computational cost.

Another possibility is that self-consistent electron heating reduces variability. In Appendix B we consider a set of such models from Dexter et al. (2020). These models show a variability excess that is similar to the fiducial models.

The $R_{\rm high}$ prescription contains a parameter $R_{\rm low}$ that determines the ion−electron temperature ratio at low β, which we have consistently set to 1. Increasing $R_{\rm low}$ models the effect of rapid cooling in low-β regions. In Appendix B we consider $R_{\rm low}$ up to 10 in a sample of four models and show that it does not systematically reduce variability.

The variability excess might also be caused by numerical inaccuracies in radiative transfer, truncation error in the GRMHD integrations (limited resolution), limited simulation duration, or misspecification of the adiabatic index. These are considered in Appendices A and B. We find no evidence that any of these effects is producing the variability excess. A more extensive study on how these effects could alter the variability is warranted but outside the scope of this paper.

5.4. Best-bet Region without Variability

From the preceding discussion it is possible that a combination of extended flux, viscosity, cooling, and numerical limitations affects model variability enough to compromise the variability constraints. Notice that a relatively small change in variability, ∼30% in M_3, is sufficient to promote many of the MAD models from failing to passing. If we exclude the variability constraints, then, which models are favored?

Figure 15 (which also appears as Figure 33 in Appendix C) shows the result of applying all constraints except structural and flux variability to the fiducial models. Most negative-spin models are ruled out, and two MAD, positive-spin, low-inclination, large-$R_{\rm high}$ models pass all constraints for KHARMA and BHAC (the H−AMR simulations do not include $i = 30°$). Nearby models in parameter space are close to passing in the sense that they pass for one or more of KHARMA, BHAC, and H−AMR.[163]

We will call the green region of parameter space in Figure 15 the *best-bet region*. The models in this part of parameter space perform well and explain nearly all the data.

Given the uncertainty associated with the variability excess and the possibility of missing physical ingredients in the model, the existence of the best-bet region cannot be regarded as evidence that Sgr A* has positive spin and low inclination. Given the large and uncertain set of constraints, however, it is remarkable that *any* models perform as well as these do. The models in the best-bet region merit additional analysis, and it is interesting to ask what they predict for future observations.

5.5. Fiducial Model Accretion Rate and Outflow Power

What is the accretion rate in Sgr A*? The time-averaged

$$\dot{M} \equiv \frac{1}{\Delta t}\int dt \int d\theta d\phi\ \sqrt{-g}\,(-\rho u^r) \tag{20}$$

[163] If we use the more permissive 86 GHz size constraint from Issaoun et al. (2019), we find two new best-bet models: one MAD model close to the best-bet region at $a_* = 0.94$, $i = 10°$, and $R_{\rm high} = 160$; and one SANE model at $a_* = 0.5$, $i = 30°$, and $R_{\rm high} = 10$.

at the event horizon; the quantity in parentheses is the inward rest-mass flux density.

Figure 16 (top panels) shows $\dot{M}$ in solar masses per year for the KHARMA and BHAC thermal models. The accretion rates follow immediately once the models are normalized so that $\langle F_{230}\rangle = 2.4$ Jy.

MAD models accrete at 10^{-9} to $10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$, while SANE models have a broader range, 10^{-9} to $10^{-6}\,M_\odot\,\mathrm{yr}^{-1}$. $\dot{M}$ is an increasing function of $R_{\rm high}$, which is a result of the thermal synchrotron emissivity increasing with increasing electron temperature, so that at fixed flux density lower electron temperature models (higher $R_{\rm high}$) have higher $\dot{M}$. SANE models exhibit a stronger dependence on $R_{\rm high}$ than MAD models. The emission in the SANE models is predominantly from regions with large β, where the electron temperature is regulated by $R_{\rm high}$. MAD models produce more of their emission in regions with $\beta \sim 1$, where $R_{\rm high}$ only has a weak effect on electron temperature (see M87* Paper V, Figure 4). For both SANE and MAD models, $\dot{M}$ decreases with increasing spin. This follows from the dependence of temperature on spin shown in Figure 2—higher spin models have higher gas temperature, which in the $R_{\rm high}$ model implies higher electron temperature.

Retrograde SANE, $R_{\rm high} = 40$ and 160 models produce the largest $\dot{M} \sim 10^{-6}\,M_\odot\,\mathrm{yr}^{-1}$. These models have a high midplane density of cool electrons that, in KHARMA models, overproduce X-ray emission through bremsstrahlung and are therefore ruled out. Critical beta models have accretion rates that lie between the fiducial model values for $R_{\rm high} = 10$ and $R_{\rm high} = 40$ for the selected critical beta parameter values ($f = 0.5$, $\beta_{\rm crit} = 1$).

How do our $\dot{M}$ compare with earlier estimates? Linear polarization and Faraday rotation measurements at millimeter and submillimeter (submm) wavelengths (Agol 2000; Quataert & Gruzinov 2000; Bower et al. 2003; Macquart et al. 2006; Marrone et al. 2006a, 2006b) and X-ray emission (Baganoff et al. 2003; Wang et al. 2013) combined with semianalytic models predict $\dot{M} \sim 10^{-9}$ to $10^{-7}\,M_\odot\,\mathrm{yr}^{-1}$. The broad range of values is due to the differences in regions of radio emission in the theoretical models that are considered (ADAFs: Narayan et al. 1998; Yuan et al. 2003; jet models: Falcke et al. 1993; Falcke & Markoff 2000; ADAF+jet models: Yuan et al. 2002). All our MAD model $\dot{M}$ fall within the range of these historical observational estimates. In contrast, almost all SANE models with larger $R_{\rm high}$ parameter (except SANE $a_* = 0.94$, which is one of our best models) have an $\dot{M}$ that is inconsistent with earlier estimates.

All fiducial models produce outflows in the polar regions. In many cases the outflows can be divided into a slower, denser disk wind and a relativistic, high-σ Poynting jet. The outflows have a power that is comparable to or larger than the bolometric luminosity. What is the outflow power $P_{\rm out}$?

First, we must define $P_{\rm out}$. There are a number of competing definitions in the literature; we set

$$P_{\rm out} \equiv \frac{1}{\Delta t}\int dt \int d\phi \int_{\rm poles} d\theta \sqrt{-g}\,(-T^r_t - \rho u^r), \qquad (21)$$

where "poles" indicates $\theta < 1$ rad or $\theta > (\pi - 1)$rad. We include only those θ where the time- and azimuth-averaged energy flux is outward. The integral is evaluated at $r = 100\,GM/c^2$. $P_{\rm out}$ includes power in the relativistic Poynting jet, if present, and in the slower, denser disk wind.

Figure 16 (bottom panels) shows $P_{\rm out}$ for the fiducial KHARMA and BHAC models. As expected, the outflow power increases with the magnitude of black hole spin. We find that SANE models have 10^{35} erg s^{-1} $\lesssim P_{\rm out} \lesssim 10^{38}$ erg s^{-1}. Evidently $P_{\rm out}$ increases with $R_{\rm high}$, as expected from the behavior of $\dot{M}$: higher-$\dot{M}$ models have stronger magnetic fields. We find that MAD models have 10^{37} erg s^{-1} $\lesssim P_{\rm out} \lesssim 10^{39}$ erg s^{-1} and are, on average, more powerful and less sensitive to $R_{\rm high}$.

Many high-spin or large-$R_{\rm high}$ models have $P_{\rm out} \sim 10^{38}$ erg s^{-1}. An outflow with this power could produce dramatic observable effects in the dense interstellar medium (ISM) of the Galactic center. For instance, the X-ray transient CXOGC J174540.0 −290031, located only 0.1 pc from Sgr A* and with an estimated jet power of $\sim 10^{36}$ erg s^{-1}, produced a compact bipolar lobe at radio wavelengths with peak flux densities near 100 mJy (Bower et al. 2005). A more continuous outflow, however, might clear out a substantial volume of space, making identification of any interaction with the ISM less certain. Nevertheless, there have been a number of large-scale features that have been suggested as the result of interaction of a jet with the ISM (e.g., Li et al. 2013; Cecil et al. 2021).

5.6. Caveats and Limitations

5.6.1. Electron Distribution Function

One of the central uncertainties in modeling Sgr A* is eDF assignment. Do the surviving models have anything to say about the eDF?

In our fiducial models, thermal eDFs with equal ion and electron temperature ($R_{\rm high} = 1$) are ruled out, in most cases by more than one constraint. For MAD models this is easily explained: Comptonization is strong at $R_{\rm high} = 1$, and X-rays are overproduced. For MAD models, $\dot{M}$ and therefore the electron-scattering optical depth $\tau_{\rm es}$ are insensitive to $R_{\rm high}$. Since the amplitude of the first Compton bump is $\propto y = 16\Theta_e^2\tau_{\rm es}$, the high Θ_e at $R_{\rm high} = 1$ produces a large X-ray flux. For SANE models the situation is more complicated (see Appendix C), with m-ring width rejecting many $R_{\rm high} = 1$ models at $a_* \leqslant 0$, and 86 GHz flux and size rejecting the rest. The latter is a consequence of model electron temperature reaching a maximum at large a_* and small $R_{\rm high}$, so that optical depth is a minimum and therefore so is the 86 GHz image size.

In some fiducial SANE models the sense of the X-ray constraint is reversed: bremsstrahlung is strong and X-rays are overproduced where $R_{\rm high}$ is *large*. Again, this is easily understood: when $R_{\rm high}$ is large, the midplane electrons are cold, the accretion rates (and therefore n^2) are high, and the bremsstrahlung emissivity is large. More generally, the X-rays provide a strong constraint on the presence of dark (subrelativistic) electrons, which are otherwise undetectable in millimeter wavelength emission or absorption, although they can produce strong Faraday rotation.

We have tested a large set of nonthermal models, which have a power-law tail on the eDF. Although integration times for the nonthermal models are too short to provide strong model constraints, there are trends that emerge from the existing data.

First, the 230 GHz images are relatively insensitive to the presence of nonthermal electrons for models in which most of the nonthermal electrons are introduced in and near the outflow region. This is encouraging: the 230 GHz image is generated by

Table 7
Exploratory Models That Pass All Constraints

Code/Setup	MAD/SANE	Spin	Inc	$R_{\rm high}$	Constraint(s) Failed at 15,000 GM/c^3
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	10	80	M-ring width, M_3
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	10	160	M-ring width, M_3
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	30	160	M_3
BHAC $\epsilon = 0.05$	SANE	0.94	10	10	M_3
H-AMR $p = 4$	SANE	0.94	50	40	2.2 μm flux
H-AMR $p = 4$	MAD	0.5	50	1	M-ring diameter, M_3

Note. Exploratory models that pass all constraints when computed to 5000 GM/c^3. When extended to 15,000 GM/c^3, each model fails one or more constraints.

Table 8
Exploratory Models That Fail Only One Constraint

Code/Setup	MAD/SANE	Spin	Inc	$R_{\rm high}$	Constraint Failed
BHAC $\kappa(\sigma, \beta)$	SANE	0	70	1	86 GHz flux
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	10	40	2.2 μm flux
BHAC $\kappa(\sigma, \beta)$	SANE	0.5	10	40	M_3
BHAC $\kappa(\sigma, \beta)$	SANE	0.5	30	40	M_3
BHAC $\kappa(\sigma, \beta)$	SANE	0.5	30	80	M_3
BHAC $\kappa(\sigma, \beta)$	SANE	0.5	30	160	M_3
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	30	80	M_3
BHAC $\kappa(\sigma, \beta)$	MAD	0.5	50	160	M_3
BHAC $\kappa(\sigma, \beta)$	MAD	0.94	10	160	M_3
BHAC $\epsilon = 0.05$	SANE	0.94	30	10	86 GHz flux
BHAC $\epsilon = 0.05$	MAD	0.5	30	160	2.2 μm flux
BHAC $\epsilon = 0.05$	MAD	0.5	50	160	2.2 μm flux
BHAC $\epsilon = 0.05$	MAD	0.94	10	40	2.2 μm flux
BHAC $\epsilon = 0.10$	SANE	0.94	10	10	2.2 μm flux
BHAC $\epsilon = 0.10$	MAD	0.5	30	160	2.2 μm flux
BHAC $\epsilon = 0.10$	MAD	0.5	50	160	2.2 μm flux
BHAC $\epsilon = 0.10$	MAD	0.94	10	40	2.2 μm flux
BHAC $\epsilon = 0.20$	SANE	0.94	10	10	2.2 μm flux
BHAC $\epsilon = 0.20$	MAD	0.94	10	40	2.2 μm flux
BHAC $\epsilon = 0.20$	MAD	0.94	10	160	2.2 μm flux
BHAC $\kappa = 5$	MAD	0.94	50	1	2.2 μm flux
H-AMR 30° tilt	SANE[a]	0.94	10	160	86 GHz size

Note. Models that pass all of the constraints except for one. Since no model passes all constraints, these represent the parameters that are closest to being consistent with observations.
[a] For the tilted model, $\phi/\phi_{\rm crit} \simeq 0.8$.

electrons in an approximately thermal core of the eDF and is relatively insensitive to the behavior of the tails.

Second, as one might expect, the 2.2 μm flux density is an increasing function of nonthermal electron density. In many models (e.g., the variable efficiency models of Section 4.2.3) the addition of a power-law tail changes a thermal model that passes the 2.2 μm test into a nonthermal model that fails.

Third, it is important to understand that many of the nonthermal models we use are linked to the $R_{\rm high}$ prescription in some way. For example, the κ distribution function contains a width parameter w, and this is set using an $R_{\rm high}$-like prescription with width depending on β. Our nonthermal models are only a few points in a vast function space of possible nonthermal parameterizations, with none of the models considered allowing for an electron energy density that depends on plasma history and instantaneous plasma state.

5.6.2. Collisionless Plasma Effects

The mean free path to Coulomb scattering for particles is typically larger than or comparable to the system size in Sgr A*, rendering its plasma collisionless. The GRMHD simulations used in this work describe a collisional system, whereas a first-principles modeling of the collisionless plasma requires a fully kinetic treatment. General relativistic (radiative) kinetic simulations are crucial for dynamically probing the electron temperature, effects of nonthermal distribution functions, and pressure anisotropy and their interplay with radiation in collisionless plasma in the accretion disk and jet. While global general relativistic kinetic simulations cannot be performed with full physical separation between microscopic plasma scales (the particle's Larmor radius $r_{\rm L}$, and plasma skin depth d_e) and macroscopic scales (the gravitational radius r_g), they can achieve the right hierarchy of scales ($r_g \gg d_e \gg r_{\rm L}$) for magnetized plasmas (Chen et al. 2018; Levinson & Cerutti 2018; Parfrey et al. 2019; Chen & Yuan 2020; Crinquand et al. 2020; Kisaka et al. 2020; Bransgrove et al. 2021; Crinquand et al. 2021). Even in GRMHD, it is computationally challenging to resolve plasma heating processes powering the observed radiation in a converged manner. It is not yet feasible to resolve dissipation at the smallest scales of the turbulent cascade or the interplay between turbulence and reconnection at a similar level to that in local box simulations (Riquelme et al. 2012; Hoshino 2013, 2015; Kunz et al. 2016; Zhdankin et al. 2017; Comisso & Sironi 2018; Inchingolo et al. 2018; Zhdankin et al. 2019; Nättilä & Beloborodov 2021; Chernoglazov et al. 2021). However, Porth et al. (2019) and H. Olivares et al. (2022 in preparation) show that the global accretion dynamics (mass accretion rate, magnetic flux on the horizon, and MRI quality factor) are converging between the different simulations in this work. Kinetic processes in the (near-)collisionless plasma may increase the effective particle collision rate (see, e.g., Kunz et al. 2016). Deviations from the infinitely conducting ideal fluid approximation may alter the thermodynamics of the flow (see, e.g., Foucart et al. 2017, and Appendix C1). Some aspects of (near-)collisionless plasma dynamics can be described with nonideal effects (e.g., viscosity, resistivity, heat conduction, pressure anisotropy) in GRMHD simulations of black hole accretion (e.g., Bugli et al. 2014; Chandra et al. 2015; Foucart et al. 2016; Chandra et al. 2017; Foucart et al. 2017; Qian et al. 2018; Ripperda et al. 2019; Vourellis et al. 2019; Ripperda et al. 2020; Most & Noronha 2021; Nathanail et al. 2021; Most et al. 2021). For example, the first efforts have recently been made with high-resolution global GRMHD simulations to capture heating through magnetic reconnection in the largest current sheets in the system (Nathanail et al. 2020; Ripperda et al. 2020;

Chashkina et al. 2021; Nathanail et al. 2021; Ripperda et al. 2022).

5.6.3. Positrons

So far we have considered only ion−electron plasmas, but pairs can be produced in the 230 GHz emission region through pair discharges or through so-called pair drizzle.

The importance of pairs has been assessed using phenomenological models (Anantua et al. 2020a; Emami et al. 2021) and depends sensitively on the efficiency with which a reservoir of magnetic energy can be converted into pairs. If this efficiency is large, then pairs can substantially increase intensity in the jet region.

Production of pairs through the drizzle process is weak in Sgr A* because its luminosity is low. Wong et al. (2021; see also Mościbrodzka et al. 2011) estimate the drizzle pair density of Sgr A* to be $10^{-8}\,\mathrm{cm}^{-3}$. This is well below the Goldreich−Julian density $\sim a_* Bc^2/(4\pi eGM) \sim 10^{-2} a_* [B/(30\,\mathrm{G})]\,\mathrm{cm}^{-3}$ required to screen electric fields, suggesting that pair discharges are likely. If pair discharges serve only to raise the pair density to the Goldreich−Julian density, however, then the jet is unlikely to outshine the accretion flow, where the magnetic field strength is similar to that in the jet but the characteristic number density is $\sim 10^6\,\mathrm{cm}^{-3}$. The maximum conceivable impact of pairs on EHT observations is obtained by converting about half of the magnetic energy density into pairs with Lorentz factor ∼30 so that in Sgr A*'s ∼30 G magnetic field the emissivity of the resulting pairs peaks close to 230 GHz. Then, the pair density $\sim 10^6\,\mathrm{cm}^{-3}$ and the jet might compete with the accretion flow.

5.7. Outlook

Except for a brief discussion in Section 5.1, our analysis omits polarization. Future analyses will test our models against integrated polarization of Sgr A* (Goddi et al. 2021) and against polarized imaging, as was done for M87* (M87* Paper VII; M87* Paper VIII).

Our analysis also omits discussion of one of the main observational features of Sgr A*: the NIR and X-ray flares, for which there is as yet no consensus model. Our analysis is built on the notion that the NIR flares, at least, could be produced by accelerating a small fraction of the electron population into a nonthermal tail. The increase in 2.2 μm flux density with nonthermal population seen in Section 4 is consistent with this notion.

The agreement between our results on the source size and orientation and those of the GRAVITY Collaboration (Gravity Collaboration et al. 2018) also support the similarity in the spatial distribution of the electrons producing NIR flares and those responsible for the 230 GHz emission. Gravity Collaboration et al. (2018) find that the NIR flares originate from a region ≈40–50 μas from the black hole, only slightly larger than the diameter of the m-ring fit to the 230 GHz image. Moreover, the combined results of our analysis point toward a low observer inclination, again consistent with the GRAVITY results, although the GRAVITY results are based on a model with a hot spot orbiting at $\lesssim$40° from the plane of the sky.

In M87* we were able to identify a sense of rotation of the source from the asymmetry of the observed ring and the orientation of the large-scale jet. For Sgr A* we have not yet been able to identify a preferred PA for the source or measure the amplitude of source asymmetry, perhaps because it is small (Paper IV). The sparse baseline coverage from 2017 sharply limits our ability to detect asymmetry, but the 2021 observation already had better coverage, and future EHT campaigns will add even more stations. Unless the source is aligned and $i \approx 0°$, all our models (which have a definite sense of rotation) predict that the brightest point on the ring is produced by Doppler boosting and should lie on the approaching side of the accretion flow. The sense of rotation, determined either from the helicity of spiral features in the flow or from tracking rotation of bright spots, could be compared to the clockwise motion of NIR flares observed by GRAVITY (Gravity Collaboration et al. 2018).

Our analysis shows the value of simultaneous measurements. Simultaneous or near-simultaneous GMVA observations, in particular, provide a powerful constraint on the models. The eDF is a major source of uncertainty in our analysis, and since the submm and 2.2 μm flux densities are most sensitive to the eDF, future analyses should incorporate submm data and future EHT campaigns should seek near-simultaneous submm observations.

Our analysis provides some guidance for future numerical modeling of Sgr A*. It is clear that models require long integrations ($\gtrsim$30,000 GM/c^3) to provide a converged characterization of the source. We also note that there are regions in parameter space where we may not have sampled densely enough—where, for example, the model is too large and too small at adjacent points in parameter space. An example of this is the 86 GHz size measurement, which is sensitive to inclination, as seen for the goldilocks model. We found that being able to compare three simulation pipelines helped us identify numerical sensitivities and saved us from error on a number of occasions. One point raised in these comparisons is that the 2.2 μm flux density is sensitive to where emission is cut off in high-σ regions of the flow—the so-called σ_{cut} parameter. This point merits future investigation.

Throughout this work we have assumed that the mass and distance to Sgr A* are known. This assumption could be relaxed and the models checked for consistency with the stellar orbit measurements of mass and distance. Phenomenological accretion flow models are particularly well suited to this type of study since they are inexpensive to compute (e.g., Broderick et al. 2009).

6. Conclusions

We have made a first comparison of the EHT 2017 Sgr A* data to a state-of-the-art library of ideal GRMHD models. The models assume that the mass and distance to Sgr A* are known and that the central object is a black hole described by the Kerr metric. We use multiple simulation pipelines and find that, for a given model configuration, independent simulations are remarkably consistent (Appendix A).

The model parameters are as follows: whether the horizon magnetic field is strong or weak (MAD or SANE, respectively); the black hole spin a_*; and the inclination angle i between the line of sight and the accretion flow orbital angular momentum vector. The eDF also has one or more parameters. In our "fiducial" model set, run with three independent codes, the eDF is determined using the so-called R_{high} prescription (Section 2). We have also considered exploratory models with alternate eDF prescriptions and alternate initial conditions.

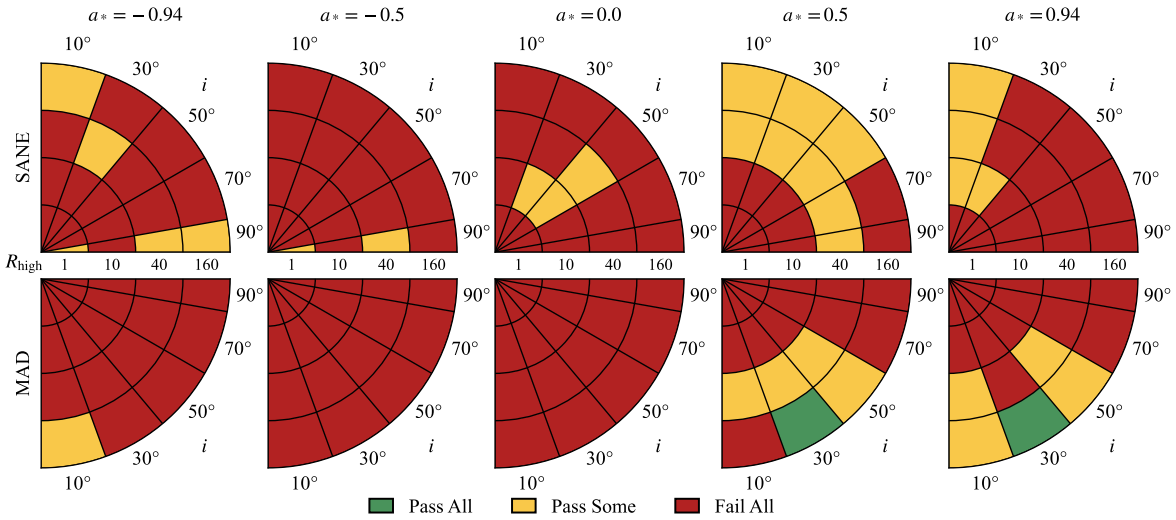

Figure 15. Combined constraints without structural or flux variability. Green indicates that the KHARMA, BHAC, and (for $i = 10°, 50°, 90°$) H-AMR models pass, yellow that one or two of the fiducial models fail, and red that all models fail.

We have selected and applied 11 heterogeneous observational constraints. Six derive directly from EHT VLBI data, two derive from 86 GHz VLBI observations with the GMVA, one from variability of the 230 GHz light curve, and one each from the 2.2 μm flux density and the X-ray luminosity.

Five structural constraints derive from EHT VLBI data. When combined, these constraints reject about 75% of our fiducial models. The EHT cut favors $a_* \geqslant 0$ and avoids edge-on ($i = 90°$) models and models with equal ion and electron temperatures ($R_{\rm high} = 1$). We are *not* able to constrain the source PA owing to sparse baseline coverage. The 2017 EHT observations are, nevertheless, quite constraining. New EHT observations with additional antennas will be even more constraining.

The strongest EHT-derived constraint is m-ring width. The physical interpretation of m-ring fits is challenging because fitting is done after the model is observed with limited baseline coverage and limited temporal sampling. Nevertheless, some interpretation is possible. For example, there is a trend toward lower width at higher inclination that eliminates many of the edge-on models. This can be understood since many edge-on models have higher peak brightness temperature than face-on models owing to Doppler boosting of emission from the approaching side of the disk. The flux density, which must average 2.4 Jy, is approximately proportional to the solid angle of the source multiplied by a typical brightness temperature, so when brightness temperature is higher, solid angle must be smaller. If the source is a ring of fixed radius, then higher brightness temperature implies a narrower ring.

Four constraints derive from non-EHT data that are contemporaneous or near contemporaneous. Combined, the non-EHT constraints reject 94% of fiducial models. The non-EHT cut favors strongly magnetized (MAD) models and eliminates most models at $i > 50°$ (consistent with interpretations of GRAVITY results; Gravity Collaboration et al. 2020b), and it also eliminates all models with equal ion and electron temperatures. These results highlight the value of continued multiwavelength monitoring of Sgr A*.

The non-EHT constraints, like the EHT-derived constraints, exhibit complicated but interpretable trends across parameter space. A full discussion of fiducial model trends will be explored in later papers, but as an example consider the 86 GHz size constraint. At $R_{\rm high} = 1$, the mean 86 GHz FWHM of SANE models *decreases* as inclination increases. The origin of this trend is very similar to the origin of the trend in m-ring width with inclination. At $R_{\rm high} = 1$ the electrons are hot and the source is optically thin. The peak brightness temperature of edge-on models is higher than that of face-on models owing to Doppler boosting of emission from the approaching side of the disk, and thus at fixed flux density the source becomes smaller as inclination increases. At $R_{\rm high} = 160$, by contrast, the trend is reversed and the mean 86 GHz FWHM of SANE models *increases* as inclination increases. In $R_{\rm high} = 160$ SANE models the electrons are cool in the disk midplane and most emission arises along the walls of the jet (see Figure 4 of M87* Paper V). The equatorial plane is relatively opaque and Doppler-boosted emission is hidden. In face-on models one is looking down the jet and the source appears relatively small; in edge-on models the jet is extended perpendicular to the line of sight and the source appears larger. Interestingly, MAD models exhibit only weak trends in 86 GHz size with inclination, in part because MAD models are more optically thin than SANE models and also because MAD models are more slowly rotating than SANE models, weakening Doppler boosting.

Sgr A* is variable but is not as variable as we expected based on our fiducial models. We have used two tests to compare the variability of models and data. One characterizes variability in the 230 GHz light curve (including simultaneous ALMA data), and the other characterizes structural variability expressed through fluctuations in the VAs. The light-curve variability is the tightest of all 11 constraints: it rejects 95% of our fiducial models. We find that strongly magnetized (MAD) models are more variable than weakly magnetized (SANE) models, and, grouped together, both SANE and MAD models are more variable than the data. The structural variability constraint measures the slope and amplitude of the power spectrum of the VA variability. Remarkably, we find that the power spectrum slope is consistent for all models, while the power spectrum amplitude is consistent for 43% of fiducial models.

The higher variability of the MAD models compared to the SANE models is a consequence of the quasi-regular magnetic flux expulsion events that are a defining feature of the MAD

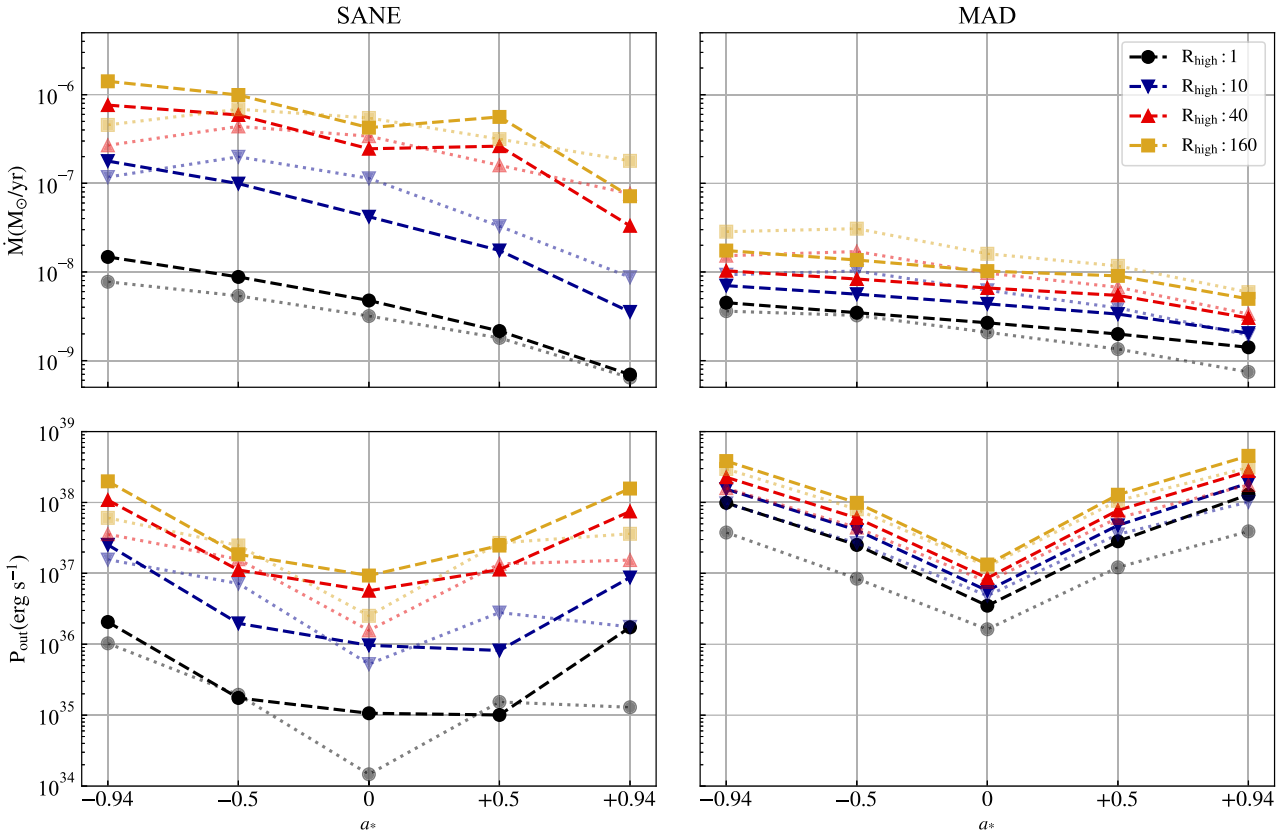

Figure 16. Comparison of the dependence on spin a_* and high end of the temperature ratio R_{high} for the accretion rate $\dot{M}$ and outflow power P_{out} in the SANE and MAD cases. Top: accretion rate $\dot{M}$ for fiducial models. Since $\dot{M}$ depends weakly on inclination i, we show only $i = 50°$. Bottom: outflow power P_{out} for fiducial models at $i = 50°$. The colors and markers vary with R_{high}. The dashed lines correspond to KHARMA thermal models, while the dotted lines indicate BHAC thermal models.

models. Magnetic flux through the horizon builds up until it exceeds a threshold and then escapes over a few dynamical times through the surrounding accretion flow. As flux is escaping, strongly magnetized low-density regions push aside plasma in the region close to the black hole that produces most of the 230 GHz emission, and this produces large fluctuations in the 230 GHz flux density. The SANE models, by contrast, exhibit relatively steady accretion through the equatorial plane.

The failure of nearly all fiducial models to match the light-curve variability is interesting. It may signal the presence of extended, slowly varying structure that is resolved out by EHT, or it may signal that future models need to incorporate collisionless effects (potentially modeled as viscosity and conductivity) or a more sophisticated treatment of electron thermodynamics including cooling. In addition, different initial geometries and polarities of the magnetic field could lead to more slowly varying structures (Nathanail et al. 2021). If, when combined, these effects were to reduce M_3 by 30%, then many MAD models would be consistent with the data.

None of the fiducial models survive the full gauntlet of 11 constraints. If we set aside both variability constraints, however, there two fiducial models that pass the remaining nine constraints in all simulation pipelines (a few more survive in one model set but not the other). These models in the "best-bet region" are strongly magnetized (MAD) and have $R_{\mathrm{high}} = 160$, positive spin, and low inclination, with (a_*, i) = (0.5, 30°) and (0.94, 30°). They have accretion rates $\dot{M} = (5.2\text{–}9.5) \times 10^{-9}\ M_\odot\,\mathrm{yr}^{-1}$, which are consistent with earlier estimates and overlap with accretion rates in wind-fed models, $\sim 10^{-8}\ M_\odot\,\mathrm{yr}^{-1}$ (Ressler et al. 2020b). The $a_* = 0.5$ MAD with $i = 30°$ model is presented in Figure 17 as snapshots and in Figure 18 as a weighted average, a convolved average, and a reconstructed average image from synthetic data. One of the reconstructed average images from the 2017 EHT observations is shown in the rightmost panel in Figure 18 for comparison.

We produced synthetic SEDs, and therefore bolometric luminosities L_{bol}, for all fiducial models. Typically L_{bol} is dominated by a synchrotron bump in the submm, and for the best-bet region it is $(6.8\text{–}9.8) \times 10^{35}$ erg s^{-1}; the corresponding radiative efficiency $L_{\mathrm{bol}}/(\dot{M}c^2)$ is $(1.3\text{–}3.0) \times 10^{-3}$. The maximum radiative efficiency over the entire fiducial model set is 0.08 (for a MAD, $a_* = 0.94$, $R_{\mathrm{high}} = 1$ model), which is necessary but not sufficient to justify our neglect of radiative cooling in the GRMHD evolution.

All our fiducial models produce bipolar outflows, and for each we measured the outflow power P_{out}, defined in Section 5. Consistent with earlier work we find that outflow power is higher for strongly magnetized (MAD) models than for comparable weakly magnetized (SANE) models and increases by more than an order of magnitude from $a_* = 0$ to $|a_*| = 0.94$. For models in the best-bet region, $P_{\mathrm{out}} = (1.3\text{–}4.8) \times 10^{38}$ erg s^{-1}, corresponding to an outflow efficiency $P_{\mathrm{out}}/(\dot{M}c^2)$ of 0.25–1.6. Such large outflow efficiency is only possible if energy is extracted from a spinning black hole via the mechanism proposed by Blandford & Znajek (1977). It is an open question how these powerful outflows might interact with incoming gas in a self-consistent